

Grado en Ingeniería de Tecnologías Industriales

2017 – 2018

Trabajo Fin de Grado

“Implementación de una Herramienta para el Análisis de los Patrones de Movimiento en Páginas Web”

Julián Brem Vilas

Tutor

Rubén Cuevas Rumin

Leganés, 2018

RESUMEN

El mundo virtual está lleno de usuarios, sin embargo, no siempre se tiene claro quién está detrás. En este trabajo se busca identificar mediante el uso de patrones de movimiento si el usuario es un ser humano o un robot.

En primer lugar crearemos una Página Web, que nos servirá como base de este proyecto. A continuación implementaremos una herramienta, la cual almacenará en una base de datos, los movimientos que vaya realizando el usuario. Más tarde generaremos un “bot” para que navegue por la página digital y así poder registrar también su comportamiento.

Con los datos obtenidos, se buscará un patrón de movimiento que pueda diferenciar el comportamiento de un robot y un ser humano. Como se podrá ver, los resultados muestran una clara diferencia entre los dos grupos, usuario y robot, debido al comportamiento de cada uno de ellos.

Palabras Clave: Ser humano o robot, patrones de movimiento, Página Web.

DEDICATORIA

Al finalizar este trabajo de fin de grado me gustaría expresar mi agradecimiento a todas esas personas que me han ayudado en la elaboración del mismo. Particularmente agradecer la ayuda y el estímulo que me ha otorgado mi familia, mención especial a mis padres y hermanos. Así mismo me gustaría agradecer a Patricia Callejo la oportunidad que me ofreció de realizar este trabajo bajo su dirección.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1. Motivación del trabajo	1
1.2. Objetivos	2
2. ESTUDIO DE LAS TECNOLOGÍAS	3
2.1. Herramientas de detección de web-bots	3
2.2. Lenguajes de programación	6
2.3. Bases de datos	7
3. DISEÑO DE LA HERRAMIENTA Y DEL BOT	10
3.1. Diseño y requisitos	10
3.2. FRONT END	14
3.3. BACK END	17
3.4. Usuarios reales y bot	21
3.4.1. Recolección de los usuarios reales	21
3.4.2. Creación del bot	22
4. ANÁLISIS DE LOS DATOS OBTENIDOS	24
5. MARCO REGULADOR	35
5.1. Directiva Europea de protección de datos.	35
5.2. Regulación en España	36
6. ENTORNO SOCIO-ECONÓMICO	37
6.1. Impacto económico	37
6.2. Impacto social	38
7. PLANIFICACIÓN DEL PROYECTO	39
7.1. Planificación	39
8. CONCLUSIONES	41
I. ABSTRACT	44
I.I. Motivation of the Project	44
I.II. Objectives	45
I.III. Analysis of results	45
I.IV. Conclusions	46
II. REFERENCIAS	47

ÍNDICE DE ILUSTRACIONES

Ilustración 3.1.1.	Código Html	11
Ilustración 3.1.2.	Código Css	12
Ilustración 3.1.3.	Código Php	13
Ilustración 3.1.4.	Código JavaScript.....	14
Ilustración 3.1.5.	Código JQuery	14
Ilustración 3.2.1.	Páginas: Inicio y Pádel	15
Ilustración 3.2.2.	Página: Página web.....	15
Ilustración 3.2.3.	Página: Formulario	16
Ilustración 3.2.4.	Menú Login	17
Ilustración 3.3.1.	Documentos en una colección de MongoDB	19
Ilustración 3.3.2.	Estructura documento anónimo	20
Ilustración 3.3.3.	Estructura documento identificados	21
Ilustración 4.1.	Inicio: Usuarios anónimos	25
Ilustración 4.2.	Página web: Usuarios anónimos.....	25
Ilustración 4.3.	Programa: Usuarios anónimos.....	26
Ilustración 4.4.	Pádel: Usuarios anónimos	26
Ilustración 4.5.	Estiramientos: Usuarios registrados	27
Ilustración 4.6.	Lesiones: Usuarios registrados	28
Ilustración 4.7.	Inicio: Usuarios registrados	28
Ilustración 4.8.	Página web: Usuarios registrados.....	29
Ilustración 4.9.	Estiramientos: Robot	31
Ilustración 4.10.	Lesiones: Robot	32
Ilustración 4.11.	Inicio: Robot.....	32
Ilustración 4.12.	Página web: Robot.....	33
Ilustración 4.13.	Diferencias entre ser humano y robot.....	34
Ilustración 7.1.1.	Diagrama de Gannt.....	39

ÍNDICE DE TABLAS

Tabla 2.3.1.	Diferencias entre bases de datos	8
Tabla 4.1.	Tiempos de visualización de los usuarios	30
Tabla 4.2.	Tiempos de visualización del autómata.....	30
Tabla 8.1.	Recursos Materiales.....	40
Tabla 8.2.	Recursos Humanos	40

1. INTRODUCCIÓN

1.1. Motivación del trabajo

A lo largo de estos años hemos visto un notable aumento de la seguridad a la hora de verificar los datos del usuario para iniciar sesión en una página web. En la actualidad, vemos en muchas ocasiones, unas casillas en las que aparecen cuadrículas con diferentes imágenes para que identifiquemos en cuales sale una figura determinada o por otro lado también podemos ver recuadros, en los que aparecen un grupo de caracteres tachados y difíciles de leer y nos piden que repitamos la secuencia (Captcha). Este tipo de herramienta se ha creado para poder verificar si el sujeto que está accediendo a la página web es en efecto un ser humano, ya que, en la actualidad, los “bots” todavía no son capaces de pasar estas pruebas.

Todo esto me llevó a plantearme la siguiente pregunta: ¿Hay alguna otra forma de verificar si el sujeto que está accediendo a mi página digital es un ser humano?

Como todos sabemos, al navegar en una página web generamos diferentes movimientos ya sea al hacer zoom, al copiar algún fragmento, al hacer click o al movernos con el ratón. Todos estos movimientos nos están dando datos de quién se encuentra detrás de ese navegador. Actualmente, en las páginas web se está utilizando una herramienta llamada “Cookie” que se encarga de archivar pequeñas cantidades de datos que se intercambian entre el emisor (servidor) y el receptor (navegador). Su principal función es identificar al usuario que está accediendo a la página y ofrecer el contenido más apropiado según sus hábitos de navegación que han quedado grabados en esos pequeños archivos. Es decir, cuando entramos por primera vez en un sitio web, se guarda una cookie con una pequeña cantidad de información, que será utilizada la siguiente vez que se acceda a esa página digital para proporcionar al usuario una visita lo más personalizada posible, sin embargo, ésta no es capaz de diferenciar quien está detrás de cada usuario.

No obstante, hay que mencionar que en la actualidad más del 45% del tráfico web esta generado por los bots, los cuales, en muchas ocasiones, consiguen falsear estadísticas, manipular los anuncios estafando la publicidad digital, falsear los datos sobre los

usuarios interesados en adquirir la publicidad, etc. Todo esto genera un problema importante, ya que tanto los anunciantes como los consumidores acaban perdiendo la confianza en la publicidad digital, la cual es uno de los motores que más cantidad de dinero mueven a diario en internet

El resolver esa incógnita me ha llevado a desarrollar este trabajo de fin de grado ya que me daba la oportunidad de crear una herramienta diferente, con la cual podemos almacenar los movimientos del ratón y del teclado y mediante un análisis de los mismos determinar, en definitiva, un patrón que sea capaz de diferenciar el comportamiento de una persona y un robot.

1.2. Objetivos

El objetivo principal para este proyecto será la implementación de una herramienta que analice el comportamiento de diferentes usuarios (seres humanos y robots), basándonos en los movimientos que se lleven a cabo y determinar así un patrón lógico, que nos permita identificar quien está detrás de cada visita a la página web.

Antes de generar esta herramienta tendremos que diseñar primero una Página Web, con el fin de crear una plataforma que pueda ser visitada por diferentes personas, la cual deberá estar vinculada con una base de datos y de esta forma podremos registrar los diferentes comportamientos de los usuarios. Y por último crearemos un ‘bot’ con la finalidad de generar un usuario falso, que simule el comportamiento de un robot.

El análisis de todos los resultados nos llevará a responder esa incógnita planteada al principio: ¿Es un robot o un ser humano?

2. ESTUDIO DE LAS TECNOLOGÍAS

Antes de empezar, vamos a explicar brevemente el significado que tiene el término de “página web”:

La página web o también conocida como página electrónica, digital o ciberpágina es un documento virtual al cual se accede mediante un navegador web (Google Chrome, Internet Explorer, Mozilla Firefox, Internet Explorer, etc.). Este documento está formado por diferentes partes, está capacitado para contener texto, programas, sonido, video, imágenes, etc. definidos con HTML y XHTML y está adaptado para la llamada World Wide Web (www).

Normalmente estas páginas electrónicas están almacenadas en equipos locales o en un servidor web remoto. El servidor tiene dos opciones o bien restringe el acceso a la página a redes privadas (intranet corporativa) o bien puede publicar las páginas en la World Wide Web y hacerlas visibles a todos los que dispongan de un navegador web.

Partiendo de esta base en este proyecto el desarrollo de la página digital es un aspecto fundamental, ya que sin él no podríamos recoger los datos de los usuarios y no podríamos llevar a cabo su análisis.

2.1. Herramientas de detección de web-bots

Con el aumento de las redes sociales y los diferentes servicios web nos encontramos cada vez con más robots que se están haciendo pasar por un ser humano y esto, a menudo, nos crea inconvenientes a la hora de gestionar un sitio web.

Por ello se han creado algunas herramientas⁽¹⁾ que son capaces de rastrear los movimientos de los usuarios como: Click Tale⁽²⁾, Hotjar⁽³⁾, CrazyEgg⁽⁴⁾, improvely⁽⁵⁾, entre otros. Estos programas son capaces de generar un mapa de calor para visualizar los distintos movimientos del ratón, crean grabaciones de la navegación en el sitio web, rastrean los clicks, etc. Cada una de estas aplicaciones tiene un objetivo y un enfoque diferente. Algunas interactúan con la publicidad que se muestra centrándose en los factores SEO (Search Engine Optimization), otras se fijan en los movimientos del usuario para determinar qué es lo más atractivo de la página y así mejorar su estrategia de marketing creando una visita única y personalizada. Sin embargo estas herramientas

no pueden definir si un usuario es persona o robot. Haremos a continuación una breve reseña de alguna de ellas.

Click Tale es un programa que nos facilita el análisis web y nos muestra los comportamientos de los usuarios. Se caracteriza por ser fácil de implementar y muy útil, ya que nos enseña mapas de calor, diferenciando entre las tres funcionalidades que nos ofrece un ratón (mouse move, scroll, clicks) y quedando registrada la actividad del usuario. Embudos de conversión para detectar sitios de mejora en la web y conseguir así una mejor experiencia de navegación para el usuario, detectando problemas y posibles mejoras, grabaciones de su navegación pasando por el tiempo que ha transcurrido mientras estaba en la página hasta ver desde que enlace ha llegado a ella.

Hotjar es una herramienta que te ayuda a mejorar la experiencia de los usuarios y elimina todo aquello que no sea relevante, genera encuestas contextuales para determinar posibles fallos y mejoras en la web al estilo de Qualaroo, utiliza embudos de conversión para detectar los puntos débiles de la web, mapas de calor para el comportamiento del usuario e incluso analiza formularios abandonados para detectar el motivo por el cual han quedado obsoletos.

Crazy Egg es otro programa de optimización web que monitoriza la actividad de las diferentes visitas en una página web. Está destinada, más que nada, para perfeccionar la usabilidad de nuestra web, es decir, nos muestra aquellas zonas que no son relevantes para el usuario o a las que nunca ha llegado, para hacerlas más visibles y poder reestructurar la web, no obstante, todos los análisis que nos ofrece Crazy Egg están basados en los clicks que generan los usuarios, con los cuales genera mapas de calor, mapas de scroll, etc.

Improvvelly es otro software que se encarga del estudio y el análisis del comportamiento y el rendimiento de las campañas de publicidad y de marketing y tiene un sistema de seguridad para detectar y bloquear posibles intrusiones fraudulentas en la web. Es una herramienta perfecta para analizar el comportamiento de las inversiones en marketing, estudiando las reacciones que tienen los diferentes usuarios frente a las campañas publicitarias que hemos implementado en nuestra web y sacarles el mayor partido.

Todas ellas son, en definitiva, un conjunto de operaciones que nos dan mucha información sobre el comportamiento del usuario, sin embargo, como bien hemos mencionado anteriormente, no es capaz de definir la autenticidad del usuario.

Para hacer frente a esa farsa de usuarios, la mayoría de sitios web utilizan una herramienta llamada Captcha⁽⁶⁾ (Completely Automated Public Turing test to tell Computers and Humans) que aparece al iniciar la sesión o al registrarnos. Este Test ha tenido algunas variaciones con el paso del tiempo y actualmente se trata de una herramienta invisible que solo aparece cuando el programa no está seguro de la identificación del usuario. En definitiva se basa en una prueba desafío - respuesta inspirada en el test de Turing.

El test de Turing⁽⁷⁾ es un método con el cual se intenta determinar si una máquina es capaz de pensar. Se basa en un juego de imitación en el cual hay tres participantes. Un interrogador, un hombre y una mujer. El juego consiste en que el interrogador está apartado de los otros dos jugadores y tiene que identificar mediante una serie de preguntas escritas quién es la mujer y quién es el hombre, utilizando un lenguaje común para todos. Por otro lado, los otros dos participantes tienen que tratar de convencer al interrogador de que son la mujer.

La alternativa que propuso Turing consiste en sustituir a uno de los interrogados por un autómeta el cual podría verse en dos escenarios:

- En el primer escenario se cambiará el hombre por el autómeta y tendría que ser capaz de aparentar ser una mujer, para despistar al interrogador.
- En el segundo escenario se sustituiría a la mujer por el autómeta y tanto el hombre como la computadora deberán imitar las posibles respuestas de una mujer para descolocar al interrogador.

No obstante la idea principal de Turing no era descubrir el sexo de los jugadores sino, poder determinar la identidad de los participantes, por lo que decidió centrar el juego en un test de habilidades para poder distinguir entre un ser humano y un ordenador. Por otro lado para que el estudio tenga validez y se considere que el autómeta tiene la capacidad de pensar como un ser humano, tiene que engañar al interrogador un número significativo de veces.

En definitiva el test de Turing somete a la máquina a una prueba y de esta forma se ve la habilidad que tiene el robot de reproducir un comportamiento igual o similar a la de un ser humano, sin embargo, Captcha está controlado por una máquina y la prueba la realiza un ser humano, por lo que en realidad se trata del test de Turing a la inversa.

Este proyecto se ha inspirado en la idea del Captcha y en las herramientas para visualizar el comportamiento de los usuarios, sin embargo nosotros no realizaremos ninguna prueba de habilidad, ya que nos fijaremos únicamente en los movimientos del ratón y teclado en busca de un patrón de comportamiento.

2.2. Lenguajes de programación

Para la implementación de este proyecto necesitaremos un entorno de desarrollo⁽⁸⁾ web capaz de soportar varios lenguajes de programación y que se adapte a nuestras necesidades.

Actualmente para el desarrollo de páginas dinámicas es necesario programar en el lado del servidor, el cual como es ya sabido, contiene una amplia variedad de lenguajes de programación, si bien los más utilizados son PHP⁽⁹⁾, Java⁽¹⁰⁾ y ASP.Net⁽¹¹⁾.

PHP (Hypertext Pre-Processor) se caracteriza por ser un lenguaje libre, el cual permanece invisible para el usuario, ya que el código es interpretado primero por un servidor php, que manda la información resultante al usuario en formato HTML, permite generar documentos PDF, es un soporte para varias bases de datos como MySQL, Oracle, MS SQL Server, PostreSQL, entre otras, es fácil de mantener, contiene una amplia librería externa, se puede modificar con más facilidad que otros lenguajes de programación y a la hora de programar, es menos rígida para usuarios inexpertos.

Sus principales desventajas son:

- Necesita la instalación de un servidor web.
- El servidor asume todo el trabajo de la página, lo cual será ineficiente a medida que aumenten las solicitudes.
- Dificulta la modulación y la organización por capas de las aplicaciones.
- Debido a sus propiedades de conexión, entre otras, hacen que sea insegura.

Java se caracteriza por ser un lenguaje simple, dinámico, seguro, dirigido a objetos y que contiene una colección de clases que facilitan la creación de aplicaciones distribuidas. Se interpreta y compila al mismo tiempo.

Sus principales desventajas son:

- Es un lenguaje exigente y dirigido fundamentalmente a la población con experiencia en el mundo de la informática.
- Para el manejo a bajo nivel hay que utilizar métodos convencionales, lo cual limita la portabilidad.
- Puede tener problemas para conectarse con bases de datos poco comerciales.
- Alguna herramienta tiene costes adicionales.
- Tiene varios tipos de soporte técnico para la misma herramienta, que dificultan el análisis de la mejor opción.

ASP.NET (Páginas activas de servidor) es un modelo de programación fácil, está construido sobre el Common Language Runtime, tiene una amplia variedad de idiomas, un Framework con más de 4500 clases y está orientado a objetos.

Sus principales desventajas son:

- Tiene que ejecutarse en ordenadores que tengan Windows y un servidor de Microsoft IIS.
- Los lenguajes de scripting nuevos no son tan conocidos como los antiguos.

Tras analizar los diferentes lenguajes de programación existentes, sus pros y sus contras⁽¹²⁾, se ha elegido a PHP por su versatilidad a la hora de conectarse con diferentes bases de datos, por su fácil mantenimiento, por ser un lenguaje libre, flexible y poco exigente que permanece oculto frente al usuario, lo cual hace que la programación sea confiable y segura.

2.3. Bases de datos

Otro aspecto fundamental son las bases de datos⁽¹³⁾. Hoy en día disponemos de una gran variedad de gestores de datos como MySQL, Oracle, MS SQL Server, PostgreSQL, MongoDB, Redis, etc. Todas ellas se caracterizan por almacenar y retener información y podemos englobarlas en dos grupos principales:

Bases relacionales, que contienen un conjunto de datos con una estructura bien definida (MySQL, Oracle, MS SQL Server, PostgreSQL, etc.) y las no relacionales, que tienen una estructura distribuida y flexible (MongoDB, Redis, etc.).

Antes de inclinarnos por uno u otro, debemos saber qué características debe cumplir nuestra base de datos.

A continuación podemos ver en la tabla⁽¹⁴⁾ algunas diferencias entre las bases relacionales y las no relacionales, seguido de alguna breve descripción de alguno de los gestores de datos que hemos mencionado anteriormente.

BASES DE DATOS

Relacionales	No relacionales
Estructura de datos definida y consistente	Estructura de datos distribuida
Máquinas de mayor rendimiento	Máquinas de menor rendimiento
✗	Permite escalado horizontal
Relaciona los datos de diferentes extensiones	✗
✗	Distribuye grandes cantidades de información

Tabla 2.3.1. Diferencias entre bases de datos

MySQL es uno de los gestores de datos relacionales de código abierto más populares de la red, es libre, sencillo de manejar, con un buen rendimiento, funciona sobre múltiples plataformas, es seguro y soporta grandes bases de datos.

Oracle es considerado un sistema de base de datos relacionales muy completo, debido a su estabilidad, a que soporta una gran variedad de datos, a su soporte de transacciones, a su escalabilidad y su soporte multiplataforma sin olvidarnos de la gran cantidad de herramientas que existen para su administración.

MS SQL Server es un gestor de datos relacionales basado en el lenguaje Transact-SQL, está desarrollada por Microsoft, sólo es compatible con sistemas Windows, posee un entorno gráfico de administración intuitivo, permite administrar diferentes servidores desde uno mismo, tiene una buena estabilidad, escalabilidad y seguridad, sin embargo

es de pago aunque se dispone de una versión Express que permite usarlo en entornos pequeños.

MongoDB es una de las bases de datos no relacionales más populares de código abierto, es capaz de trabajar con datos con o sin una estructura definida, soporta una gran variedad de lenguajes de programación, tiene una capacidad de indexación similar a las bases relacionales con la que consigue un gran rendimiento y una gran capacidad de escalado, no soporta atomicidad y es fácil de manejar.

Redis es una base de datos no relacional que se caracteriza por almacenar los datos en tablas de hashes (clave/valor), se considera un software de código abierto, soporta una gran variedad de lenguajes de programación, tiene un rendimiento muy alto cuando se trata de datos no duraderos (usando solo memoria RAM) y no hay una diferencia notable entre la escritura y la lectura de datos.

Aunque hay una gran variedad de bases de datos, para nuestro proyecto, hemos elegido dos: una base de datos relacional MySQL para recoger los datos de registro del usuario, ya que tienen una estructura definida y el volumen de datos que se va a manejar es reducido y una base de datos no relacional MongoDB, debido a su gran rendimiento en distribución de grandes cantidades de información, que se encargará de almacenar los eventos (datos sin estructura definida) de la página web.

3. DISEÑO DE LA HERRAMIENTA Y DEL BOT

La herramienta se encargara de hacer la gestión de los datos recogidos sobre los diferentes usuarios que van a visitar nuestra web ya sean anónimos, registrados o autómatas.

A continuación veremos cómo se han diseñado y creado cada una de las partes que conforman este proyecto, empezando por la construcción de la página web y su funcionamiento y acabando con la creación de nuestro robot, para poder analizar los diferentes resultados y así concluir con el estudio.

3.1. Diseño y requisitos

Para el desarrollo de este TFG hemos tenido que desarrollar una página web sobre la que trabajaremos. Con el fin de darle un aspecto y un contenido diferente, nos hemos apoyado en diferentes lenguajes de programación utilizando como base el código php, que nos ha facilitado la interacción con las diferentes bases de datos que se encargaran de registrar e identificar a los usuarios que accedan a la página web.

En la actualidad todas las páginas web tienen un formato estándar basado en dos partes:

- Una parte estática que contiene los documentos y la estructura de la página web aplicando diferentes estilos, formas o colores.
- Una parte dinámica que se encargará de la creación de eventos, funciones, animaciones, etc.

La ciberpágina de este proyecto ha necesitado un total de cuatro lenguajes de programación, entre los que se encuentran:

 [Html^{\(15\)}](#) para el cuerpo de la web.

Es un código de programación que se utiliza para el desarrollo de páginas de Internet. Sus siglas corresponden a Hypertext Markup language, es decir lenguaje de documentos con formato de Hipertexto. Es muy simple y sirve para definir otros lenguajes que tienen que ver con el formato de los documentos. El texto en él se crea a partir de etiquetas, también llamadas tags, que permiten interconectar diversos conceptos y formatos. La estructura básica que siguen estos documentos son (Ilustración 3.1.1.):

```

<!DOCTYPE html>
<html>

<head>
  <title>Jugando con la Web</title>

  <meta charset= "UTF-8">
  <link rel="stylesheet" type="text/css" href="../fonts/estilos.css">
  <script src="../java/mousemove.js"></script>
</head>
<body>
  <div>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed di
  euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut w
  veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl
  commodo consequat.</div>
</body>
</html>

```

Ilustración 3.1.1. Código Html

<html> </html> Representa el cuerpo principal del documento Html y dentro de él se encuentra el resto.

<head> </head> Recoge todos los metadatos vinculados con el documento Html, como enlaces a hojas de estilos, scripts etc.

<title> </title> Define el nombre que aparecerá en la pestaña de navegación. Sólo admite texto, es decir lo que no sea letra no será representado.

<body> </body> Simboliza la parte principal del documento. En esta parte se pondrán todos los documentos que se quieran representar en la página electrónica Sólo puede existir un único “body” en un documento Html.

 [Css^{\(16\)}](#) para el estilo de los contenidos.

Es un lenguaje de hojas de estilos creado para controlar el aspecto o presentación de los documentos electrónicos definidos con HTML y XHTML. Sus siglas vienen de Cascading Style Sheets, es decir hojas de estilo en cascada. CSS es la mejor forma de separar los contenidos, su presentación y es imprescindible para crear páginas web complejas. Cada hoja de estilos va a determinar cómo se verán los contenidos dentro del documento HTML

Este lenguaje está basado en un conjunto de reglas: Para modificar un aspecto de la página web ya sean tablas, textos, etc. hay que identificarlos con un selector y a

continuación se determinarán los diferentes atributos situados entre corchetes como se puede ver en la siguiente imagen (Ilustración 3.1.2.):

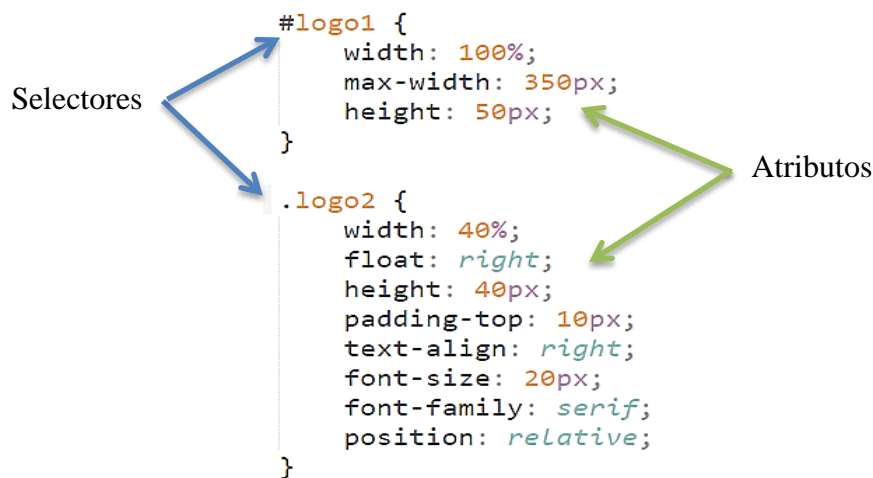


Ilustración 3.1.2. Código Css

Actualmente hay tres maneras de utilizar estas hojas de estilos:

- La primera es utilizando una hoja de estilos externa, es decir se genera un documento diferente al Html y a través de un elemento `<link> </link>` en el apartado de `<head>` se vincula esa hoja de estilos. Es la forma más inteligente y la más utilizada para agregar este tipo de documentos, ya que conseguimos separar los diferentes códigos de programación y al mismo tiempo mantenemos un orden en la creación de la web.
- la segunda forma es incluir estas líneas de código dentro del apartado `<head>` con el uso de estas etiquetas `<style> </style>`. De esta manera, también se consigue separar los diferentes códigos y mantener una limpieza a la hora de generar la web.
- La tercera manera es incluir directamente el estilo en el apartado de `<body>` al igual que antes, con el uso de las etiquetas `<style> </style>`. No obstante este uso no es recomendable, ya que a la hora de programar es aconsejable separar el contenido del aspecto visual.

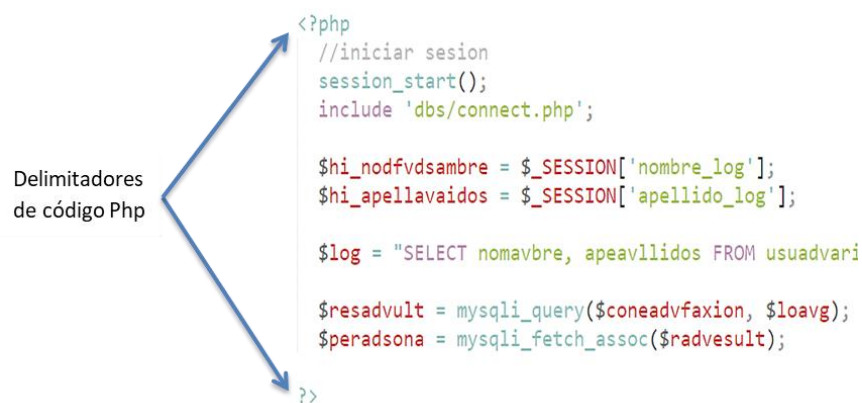
🚦 Php para el desarrollo web de contenido dinámico.

Es un lenguaje de uso general de código del lado del servidor, diseñado para el desarrollo web de contenido dinámico. El código es interpretado por un servidor web con un módulo de procesador de PHP que genera la página digital resultante.

Una característica muy importante es que se puede incorporar directamente en el documento Html, sin necesidad de utilizar un archivo externo que compile estos datos.

A diferencia del código Javascript, este lenguaje necesita que haya una petición por parte del usuario, para que el servidor pueda ejecutar el intérprete de código php y se consiga leer el documento solicitado. Por este motivo el código fuente escrito en php queda invisible para el navegador web y para el cliente, ya que el resultado que envía el intérprete al navegador está escrito en código Html. Este lenguaje, a su vez, nos facilita la conexión con diferentes bases de datos como SQL, Oracle, MySQL, Microsoft SQL Server, DB2 o MongoDB entre otras.

En la siguiente imagen (Ilustración 3.1.3.) podemos observar un ejemplo de código php.



```
<?php
//iniciar sesion
session_start();
include 'dbs/connect.php';

$hi_nodfvdsambre = $_SESSION['nombre_log'];
$hi_apellavaidos = $_SESSION['apellido_log'];

$log = "SELECT nomavbre, apeavllidos FROM usuavari

$resadvult = mysqli_query($coneadvfaxion, $loavg);
$peradsona = mysqli_fetch_assoc($radvesult);

?>
```

Ilustración 3.1.3. Código Php

✚ JavaScript y JQuery⁽¹⁷⁾ para crear animaciones, eventos en la web.

Una Web dinámica es aquella que incorpora efectos como textos que aparecen y desaparecen (popups), animaciones, eventos que se crean al pulsar botones etc. En otras palabras, para poder generar una web dinámica es necesario el uso de este tipo de lenguaje, que se caracteriza, principalmente, por estar en el lado del usuario, es decir no es necesario que se compile con antelación, ya que el propio navegador se encargará de interpretar el código en tiempo real.

La estructura de este lenguaje es similar a la usada en C y Java y al igual que con las hojas de estilo, estos también se pueden vincular al documento de Html de las tres maneras que hemos visto anteriormente, sin embargo, en este caso, utilizaremos las etiquetas `<script></script>` para referirnos a este tipo de lenguaje.

Por otro lado el código JQuery es una extensión de JavaScript. Este framework es una herramienta más que nos facilita, entre otros, la interacción con los elementos del

documento Html, desarrollar efectos, animaciones y nos ayuda a interaccionar con la tecnología AJAX en las ciberpáginas. La principal característica de este tipo de lenguaje es que nos permite cambiar el contenido de nuestras páginas sin tener que recargar la pestaña del navegador.

En las siguientes imágenes (Ilustración 3.1.4. y 3.1.5.) se pueden ver dos ejemplos sobre estos dos códigos de programación.

```

window.onload = function () {
    //Para ver el elemento que se ha pulsado y su nombre
    var ClickfbnfsHecho = document.getElementById("Element");
    if (e.srcElement) {
        var tasbsg = e.srcElement.tagName;
        var titsfbsle = e.srcElement.title;
        ClickHecho.innerHTML = "Elemento del Html = " + tag + "
        ClickHecho.style.color = "darkorange";
    } else if (e.target) {
        var tsfbsag = e.target.tagName + "&emsp; " + e.srcEl
        var tsfbsbitile = e.srcElement.title;
        ClickHecho.innerHTML = "Elemento del Html " + tag + "
        ClickHecho.style.color = "darkorange";
    }
}

```

Ilustración 3.1.4. Código JavaScript

```

$(document).on("mousedown", function(event) {
    var Htsdbvsdfbsmlpage = window.location.href;
    var cldfbsdbick = event.button;
    var butdsfbston;

    if (click == 0) {
        var XAcdfbsdbfhse = event.pageX;
        var YAcdfbsdbhse = event.pageY;
        button = "Izquierdo =" + " X: " + XAchse + "
    }
    $.ajax({
        url: '../php/dbs/connecadfbadfbt_vaadbfrfab
        type: 'post',
        datatype: 'json',
        data: {phpclick: butdsdbsfdbton, phppage:
    })
});

```

Ilustración 3.1.5. Código JQuery

3.2. FRONT END

La página web tiene un diseño sencillo a la vez que atractivo y se ha estructurado en un total de ocho sub-páginas. Dispone de un formulario para que los usuarios se puedan registrar y ver el contenido de la misma al completo.

Cada página dispone de una cabecera donde aparece el logo de la universidad, el botón de “inicio de sesión”, el cual veremos más adelante, y el título de la página (--TFG Julian Brem Vilas--) acompañado de dos imágenes.

El cuerpo principal de la página se encuentra a continuación de la barra, horizontal, del menú de navegación. Esta parte se encargará de mostrar diferentes contenidos para poder captar la atención del visitante y finalmente concluimos el diseño con un pie de página.

Por otro lado, también hay que mencionar que la barra del menú de navegación ha sido programada para que permanezca siempre visible a pesar de encontrarnos en la parte

inferior de la web, con el fin de evitar que los usuarios tengan que volver a la parte superior para cambiar de página.

A continuación veremos algunas imágenes (Ilustración 3.2.1. y 3.2.2.) para poder ver la estructura que siguen las diferentes sub-páginas, que conforman este proyecto.



Ilustración 3.2.1. Páginas: Inicio y Pádel

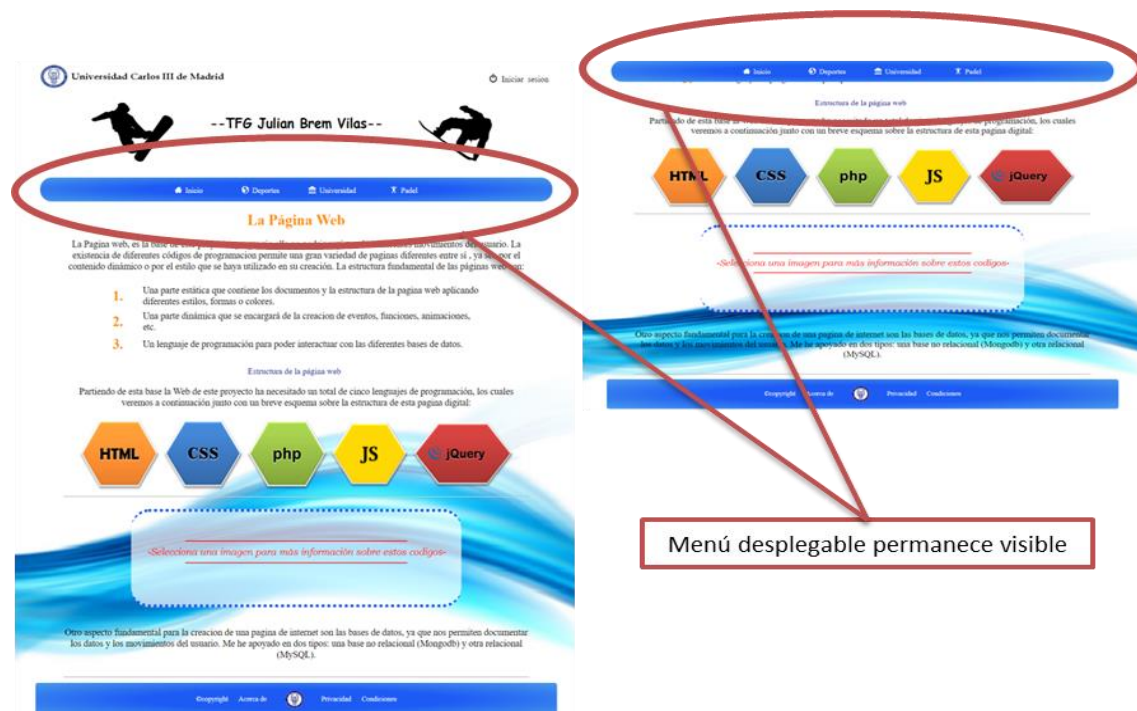


Ilustración 3.2.2. Página: Página web

En la primera imagen (Ilustración 3.2.1.) hemos detallado la estructura que siguen todas las subpáginas de este proyecto y en la segunda imagen (Ilustración 3.2.2.) se puede apreciar cómo el menú de navegación permanece visible en todo momento.

Como bien dijimos anteriormente la página dispone de un formulario, para poder iniciar sesión, y disfrutar del contenido al completo.

El formulario es bastante sencillo y está conformado por 10 campos que han de ser completados y que previamente han tenido que ser configurados, para poder ser almacenados en nuestra base de datos del tipo relacional MySQL.

El primer bloque de campos está compuesto por el nombre, los apellidos y el sexo de la persona. El segundo bloque recoge el nombre de usuario, el país de procedencia, la dirección personal. El tercer bloque se encargará de guardar el e-mail y la contraseña, la cual se tendrá que repetir para verificar que sean iguales y por último encontramos un último campo dónde recogemos los términos y condiciones de la página web.

La siguiente imagen (Ilustración 3.2.3.) nos muestra el aspecto que tiene nuestro formulario de registro.

Universidad Carlos III de Madrid

Registrate

--TFG Julian Brem Vilas--

Inicio Deportes Universidad Padel

Rellene sus datos

Nombre Apellidos

☐ Hombre ☐ Mujer

Nombre de Usuario País

Direccion*

E-mail

Contraseña Repetir contraseña

☐ Terminos y condiciones [Ver info.](#)

Enviar

*Este campo no es obligatorio

Copyright Acercas de Privacidad Condiciones

Ilustración 3.2.3. Página: Formulario

Como en la mayoría de páginas web una vez que el usuario se ha registrado correctamente podrá iniciar sesión. En nuestra página este apartado se encuentra dentro de la cabecera y se activará mediante un script situado en el botón llamado “Iniciar sesión”. Este script mostrará una pequeña pestaña, la cual está constituida por dos campos. El primer campo corresponde al e-mail y el segundo corresponde a la contraseña.

Ambos campos serán verificados y comparados con los usuarios que se hayan registrado correctamente y se encuentren dentro de nuestra base de datos relacional.

En la siguiente imagen (Ilustración 3.2.4.) vamos a ver la estructura que sigue nuestro menú de “login”.



Ilustración 3.2.4. Menú Login

3.3. BACK END

Debido a que hemos elegido php como lenguaje de programación principal, hemos tenido que habilitar un servidor web capaz de manejar nuestro entorno de desarrollo.

Para ello hemos instalado en nuestro servidor un programa llamado WampServer, el cual nos proporciona un sistema operativo (Windows), un gestor de bases de datos, un software para el servidor web (Apache) y un software de programación web (Php, Python o Perl).

Partiendo de este punto nuestro lenguaje de programación php va a pasar a ser el corazón de este proyecto, ya que actuará como enlace entre las diferentes partes que conforman este trabajo.

Actualmente existen dos métodos diferentes (método GET y método POST) para mandar información a través de php, sin embargo por motivos de seguridad y fiabilidad hemos elegido el método POST. No obstante para poder transmitir datos entre las bases de datos (BACK END) y la página web (FRONT END), es necesario combinar el método POST con los diferentes lenguajes de programación que tiene cada base de datos.

Las bases de datos no son otra cosa que una colección de información organizada, de tal manera, que puede ser interpretada por un ordenador, el cual puede acceder rápidamente a los fragmentos de datos que necesita en cada momento. En otras palabras las bases de datos son como un archivador, sin embargo, éste guarda información en formato electrónico.

Estos bancos de datos suelen estar organizados por campos, archivos y registros y como ya sabemos, existe una amplia variedad de bases de datos, que están clasificadas según el tipo de dato que se esté manejando, las necesidades que satisfagan y la utilidad que se les dará, entre otras cosas.

En particular en este proyecto hemos utilizado dos bases de datos, en concreto una base de datos del tipo relacional que almacenará los datos del formulario y otra de tipo no relacional que se encargará de almacenar los movimientos para el estudio:

1. MongoDB.

MongoDB es una base de datos NoSQL con una estructura no relacional, que está compuesta por tablas llamadas colecciones y por registros llamados documentos. Los documentos tienen una estructura de BSON, que es una representación binaria del código JSON. Al tratarse de una base de datos no relacional es más fácil de manejar, ya que no se tiene que seguir un esquema determinado. En otras palabras, los documentos que están almacenados dentro de una misma colección (tabla) no tienen que tener el mismo esquema.

En la siguiente imagen (Ilustración 3.3.1.) vemos un ejemplo sobre este comportamiento:

<pre>//Documento 1 { Nombre: "Maria", Apellidos: "Shublade", Sexo: "Mujer", Email: "amigo@ejemplo.com", Nacionalidad: [{ País: "España", Capital: "Madrid" }, { NPaís: "España", Capital: "Madrid" }] }</pre>	<pre>//Documento 2 { Nombre: "Andreas", Apellidos: "Stihl", Sexo: "Hombre", Ciudad: "Mississippi", }</pre>
---	--

Ilustración 3.3.1. Documentos en una colección de MongoDB

Como se puede observar ambos documentos siguen un esquema diferente. El documento 2 en especial contiene menos campos e incluso alguno nuevo.

Este funcionamiento es algo impensable en una base de datos relacional, ya que todos los documentos que estén dentro de una misma tabla (colección en MongoDB), deben tener el mismo esquema y el mismo número de campos.

Mongoddb es una herramienta cada vez más utilizada debido a que se puede ejecutar en máquinas con pocos recursos, soportan estructuras distribuidas (escalabilidad) y permite realizar cambios en la base de datos sin tener que detenerla.

2. Bases de datos MySQL.

Es un sistema de gestión de datos con una estructura relacional y se considera una de las bases de datos de código abierto más populares del mundo y una de las más utilizadas para el entorno del desarrollo web. Se caracteriza por tener una estructura basada en tablas y registros con diferentes campos que agrupan y separan los diferentes datos. No es recomendable para el uso de grandes volúmenes de datos.

A diferencia de las estructuras no relacionales, este tipo de bases de datos tiene un esquema determinado, lo cual nos impide ver en una misma tabla registros con un esquema diferente, es decir, todos los registros que se encuentren en una misma tabla, van a tener el mismo número de campos, por lo que a la hora de utilizar estas bases de

datos, tenemos que saber que esquema se va a utilizar y que tipo de datos vamos a ingresar.

Como bien hemos dicho anteriormente en nuestro trabajo vamos a tratar diferentes tipos de datos, por lo que a nuestra herramienta la vamos a dividir en tres partes.

La primera parte se encargará de recoger los datos de los diferentes campos que conforman el formulario, de aquellos usuarios que lo han completado exitosamente y los almacenará en una base de datos del tipo relacional (MySQL).

La segunda parte se encargará de almacenar todos los movimientos que vayan generando los usuarios en la página web, como son los eventos producidos por el ratón y el teclado, los cuales se captarán y se enviarán desde la página web al documento php por medio de jQuery. Toda esta información va ser guardada en una base de datos no relacional (MongoDB) con una determinada estructura que dependerá del tipo de usuario (documentos anónimos e identificados):

Todos los documentos anónimos se guardaran en una misma colección, sin embargo cada documento tendrá un identificador que corresponderá al User Agent y a la dirección IP del navegador que se esté utilizando para acceder a la página. El documento estará comprendido en dos campos:

- I. En el primero se almacenará la dirección http de la página en la que se encuentre el usuario.
- II. En el segundo se almacenarán los diferentes eventos del ratón y del teclado.

En la siguiente imagen (Ilustración 3.3.2.) podemos ver un ejemplo de la estructuración que seguirán los documentos anónimos.

```
"_id" : "75a244a9b--identificacion_user--5637b4f34", ← Identificador
"Inicio" : {
  "Adress" : "http://direccion de la pagina en la que se encuentre",
  "Moves" : {
    "X e Y" : [],
    "Click" : [],
    "Board Code" : [ ] } } ← Segundo campo
```

← **Primer campo**

Ilustración 3.3.2. Estructura documento anónimo

Por otro lado los usuarios registrados tendrán cada uno una colección propia donde se guardaran todos sus documentos, los cuales tendrán una estructura definida por cuatro campos:

- I. En el primero se almacenará la dirección http de la página en la que se encuentre el usuario y que servirá como identificador.
- II. En el segundo y tercer campo se almacenarán los eventos del ratón.
- III. En el cuarto se registrarán los eventos del teclado.

En la siguiente imagen (Ilustración 3.3.3) podemos ver un ejemplo de la estructuración que seguirán los documentos de los usuarios registrados.

```
"_id" : "http://direccion de la pagina en la que se encuentre",  
"X e Y" : [],  
"Click" : [],  
"Board Code" : []
```

Ilustración 3.3.3. Estructura documento identificados

Y finalmente la tercera parte se encargará de acceder a la base de datos no relacional y a continuación extraerá y representará todos los movimientos que sean relevantes para el estudio.

3.4. Usuarios reales y bot

3.4.1. Recolección de los usuarios reales

Nuestro estudio está conformado por un total de 45 usuarios anónimos y 20 usuarios registrados. La gran mayoría de los usuarios son estudiantes universitarios.

Como se puede observar hay una notable diferencia entre el número de usuarios anónimos y registrados, sin embargo no nos debemos fijar en esas cifras, ya que en algunos casos los usuarios anónimos han accedido a la web y se han dirigido directamente al formulario de registro, por lo que algunas de sus visitas no proporcionan datos que sean relevantes para nuestro estudio.

En todos los casos la información ha sido la debida mediante un mensaje de los términos y condiciones de nuestra web, por lo que nuestro proyecto cumple con las regulaciones de la nueva ley GDPR⁽¹⁸⁾ (General Data Protection Regulation).

3.4.2. Creación del bot

Para poder analizar los diferentes comportamientos atribuibles a un ser humano y a un robot vamos a elaborar un programa que sea capaz de navegar en nuestra web y de esta forma podremos registrar sus movimientos y compararlos. Como todos sabemos la principal diferencia entre una persona y un robot es, que la persona suele ser, en muchas ocasiones, impredecibles y el “bot” en cambio tienen que ser programado de antemano para que lleve a cabo una función determinada. Un ejemplo práctico serían los autómatas.

Los autómatas según la RAE (Real Academia Española) son definidos como: ‘Máquinas dotadas de un mecanismo que les permite moverse, en particular imita la figura y movimientos de un ser animado, normalmente humano’, es decir son máquinas que han sido programadas para generar un número determinado de funciones según la clase de estímulo que reciben.

Los “bots” pueden estar diseñados en cualquier lenguaje de programación. Funcionan en redes, especialmente en internet, y pueden interactuar con otros sistemas o usuarios. Pueden realizar múltiples funciones: editar textos, moderar conversaciones, responder preguntas, enviar correos electrónicos, entre muchas otras.

Hoy en día son usados para realizar diversas tareas en plataformas muy populares, como Youtube, Twitter, Facebook o Instagram.

Como bien hemos dicho anteriormente hay una gran variedad de programas, sin embargo nosotros vamos a generar un bot que simulará los movimientos del usuario:

El bot lo vamos a desarrollar mediante un lenguaje de programación llamado Java que nos permitirá realizar algunos movimientos y navegar por la página web.

Para los diferentes desplazamientos se han utilizado diferentes rectas, curvas y puntos aleatorios. Por otra parte también incluiremos el uso del teclado mediante algún comando y algunos clicks de forma no predeterminada para poder simular la actividad de una persona.

Tras un análisis detallado de las preferencias de los usuarios (anónimos y registrados), hemos seleccionado cuatro páginas por las que navegará nuestro bot, las cuales son: Inicio, Página web, Lesiones y Estiramientos. En cada una de ellas nuestro autómatas realizará movimientos diferentes apoyándose en unas ecuaciones. El tiempo de ejecución va a variar dependiendo del contenido de la página. Con el fin de poder analizar las diferencias entre el autómatas y el ser humano vamos a diseñar el bot, de tal manera, que recorra la página al completo interactuando con los diferentes scripts que haya.

Por otro lado para poder acceder de forma sencilla a los datos del bot, vamos a archivar los diferentes movimientos en una colección de un usuario registrado, la cual llamaremos “Robot”, en otras palabras, nuestro autómatas ingresará en la web como “Robot” y empezará a navegar por esas cuatro páginas que hemos mencionado anteriormente.

4. ANÁLISIS DE LOS DATOS OBTENIDOS

El estudio como bien hemos dicho anteriormente se basa en el comportamiento de los usuarios en una página web, concretamente en el rastro o la huella que van dejando (ser humano o robot) al navegar en internet.

Como todos sabemos hay una gran cantidad de bots que pueden llegar a simular el comportamiento humano, sin embargo, nuestro programa lo hemos diseñado de tal forma que simule, dentro de lo posible, la tendencia de una persona dentro de una página web.

Generalmente los seres humanos suelen tener un comportamiento similar a la hora de visualizar una página. Al principio nos quedamos en la parte superior de la web mirando el contenido y seleccionando objetos, moviendo el ratón, haciendo click, etc. Más tarde nos desplazaremos hacia la parte inferior haciendo scroll y al igual que antes realizaremos diferentes movimientos, ya sea con el ratón o el teclado, con el fin de obtener una visión global de la ciberpágina.

Este conjunto de acciones van a ser imprescindibles a la hora de programar nuestro bot, dado que queremos, que su actitud en la página electrónica se asemeje a las de un ser humano, por lo tanto nuestro autómatas, al igual que el ser humano, empezará en la parte superior y más tarde se dirigirá a la parte inferior, realizando en cada una de ellas, un conjunto de acciones.

Para el estudio tenemos una total de 45 usuarios anónimos y 20 registrados. El análisis de los datos se dividirá en las diferentes subpáginas que conforman nuestra página web. Cada grupo de usuarios dispondrá de un total de cuatro imágenes. En cada una de ellas se irán insertando los comportamientos de las diferentes visitas analizando en primer lugar los movimientos de los usuarios anónimos, en segundo lugar miraremos los movimientos de los usuarios registrados y finalmente compararemos y analizaremos los resultados con los movimientos realizados por el bot.

A continuación las siguientes imágenes nos muestran los movimientos (desplazamientos del ratón) que han realizado los diferentes usuarios en algunas de las páginas. Las primeras cuatro imágenes (Ilustración 4.1.- 4.4.) corresponden a los usuarios anónimos.

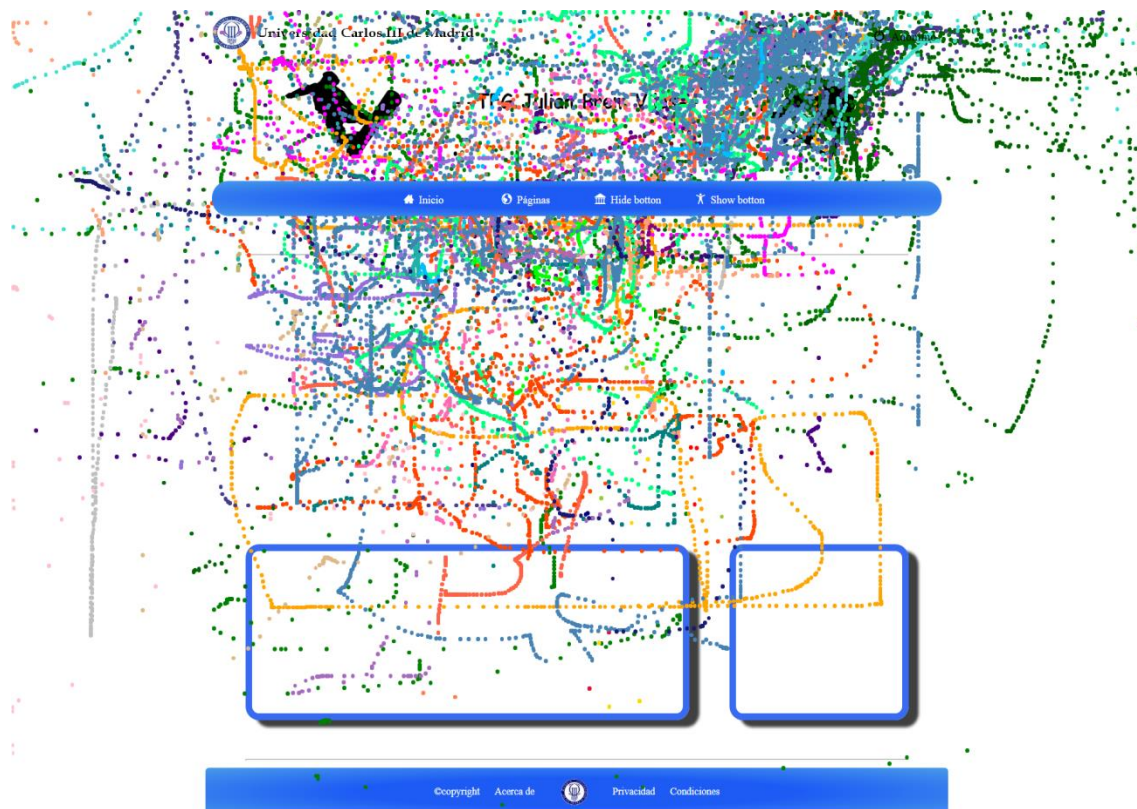


Ilustración 4.1. Inicio: Usuarios anónimos

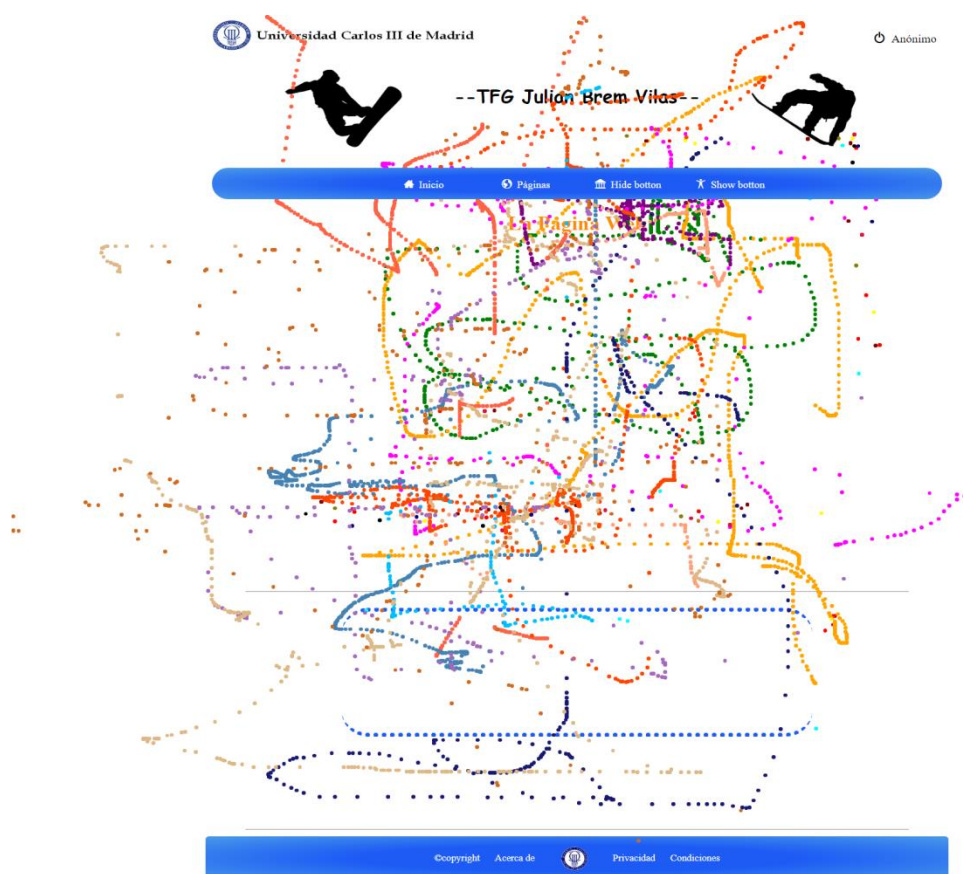


Ilustración 4.2. Página web: Usuarios anónimos

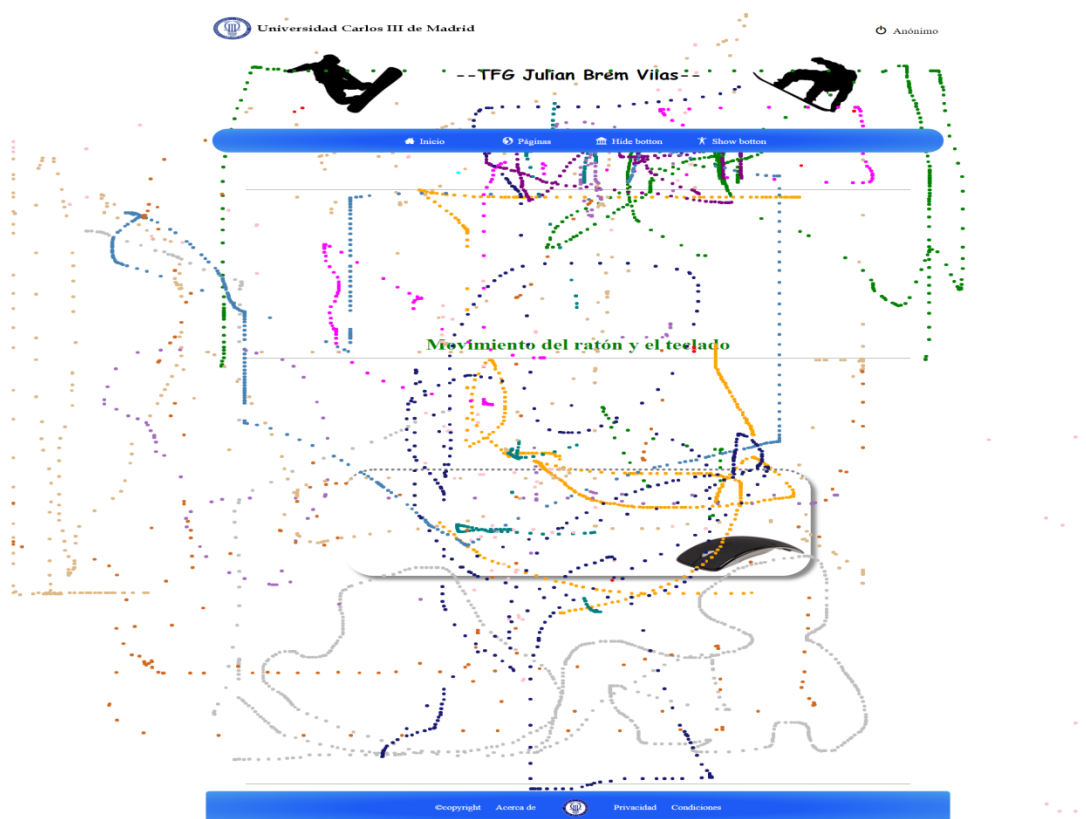


Ilustración 4.3. Programa: Usuarios anónimos

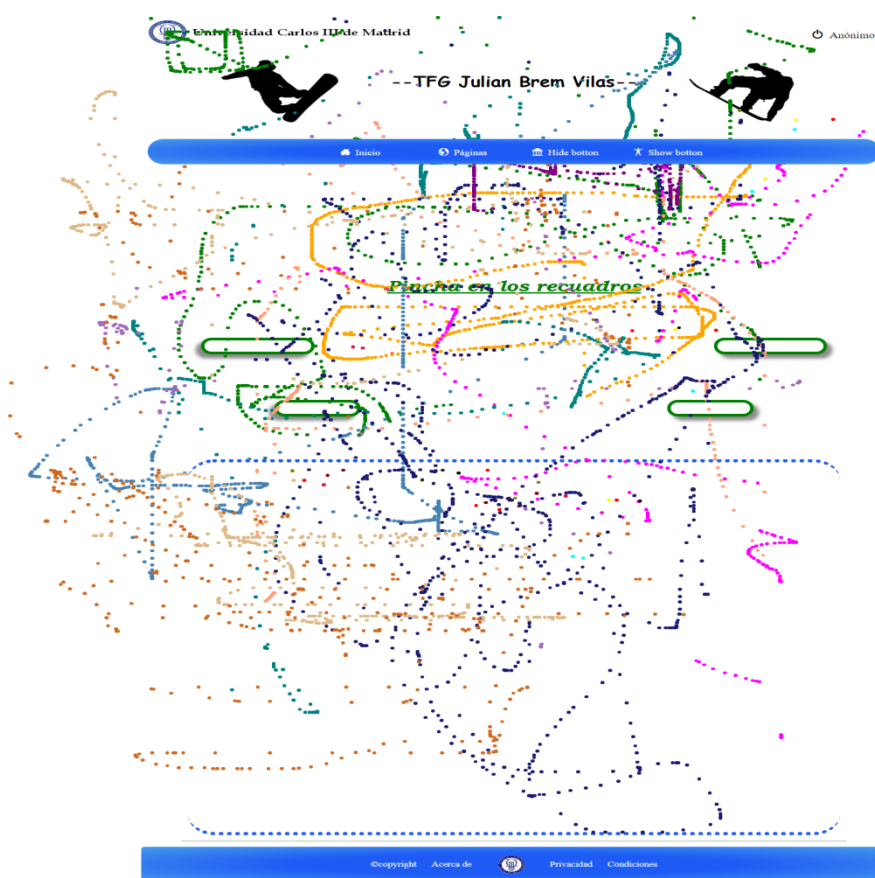


Ilustración 4.4. Pádel: Usuarios anónimos

Para el estudio de los usuarios registrados se han seleccionado dos imágenes (Ilustración 4.5.- 4.6.) que sólo ellos pueden ver (lesiones y estiramientos) y otras dos, que coinciden con los usuarios anónimos que son: inicio y página web (Ilustración 4.7.- 4.8.). El motivo por el cual se han seleccionado estas cuatro imágenes (Ilustración 4.5.- 4.8.), se debe al análisis previo sobre las preferencias de los usuarios.

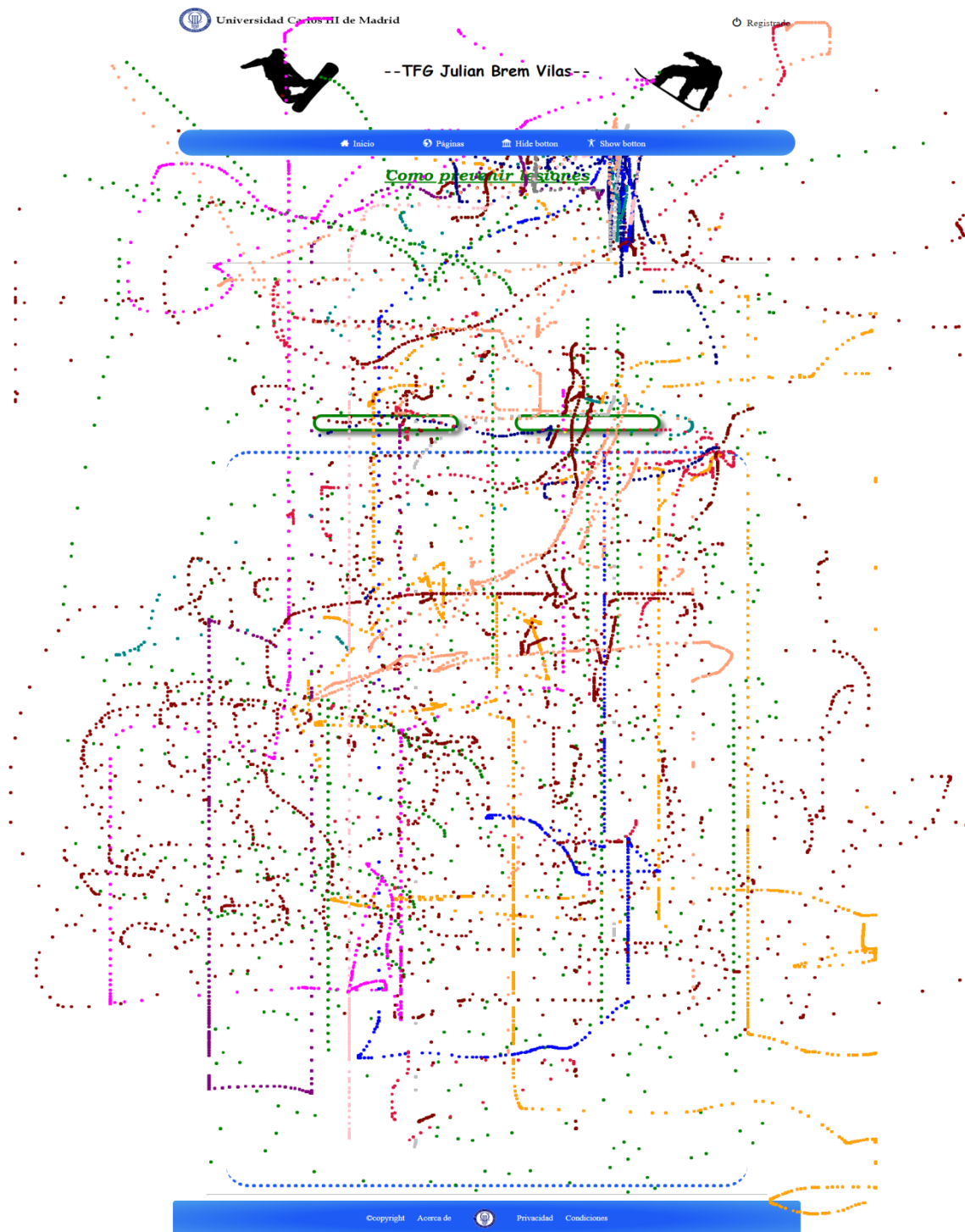


Ilustración 4.5. Estiramientos: Usuarios registrados

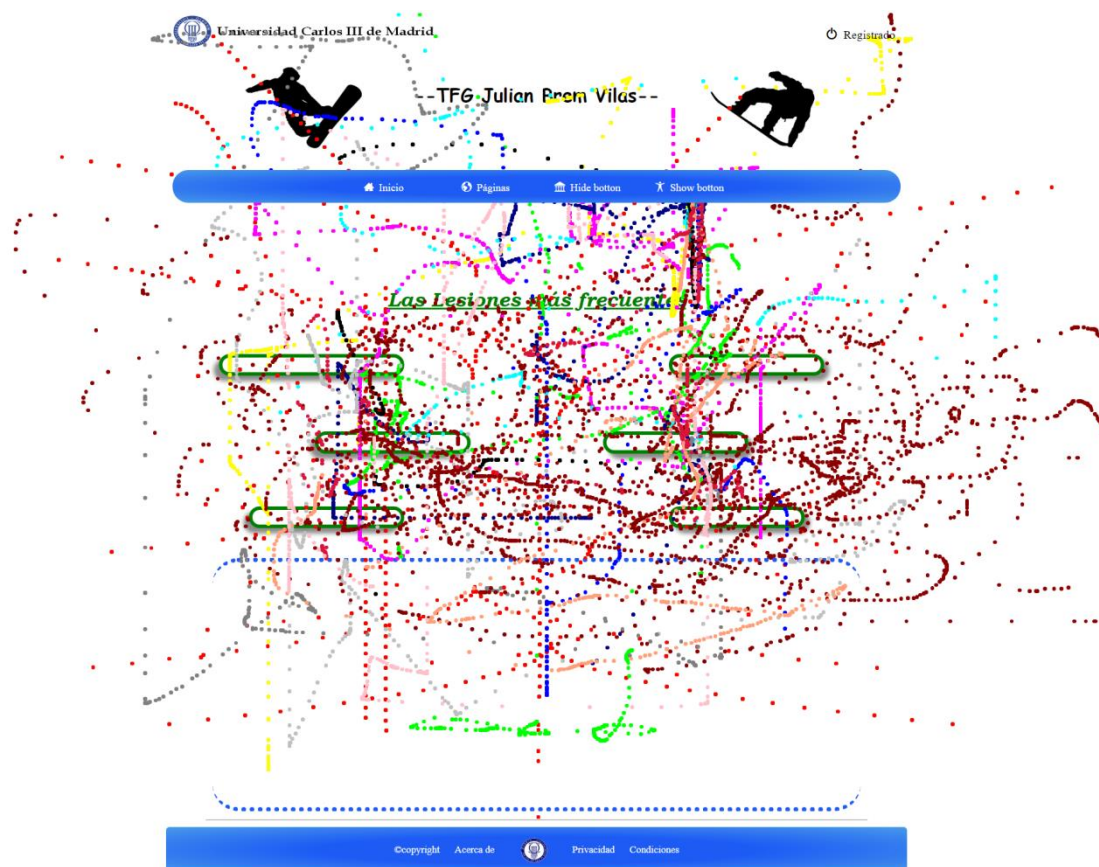


Ilustración 4.6. Lesiones: Usuarios registrados

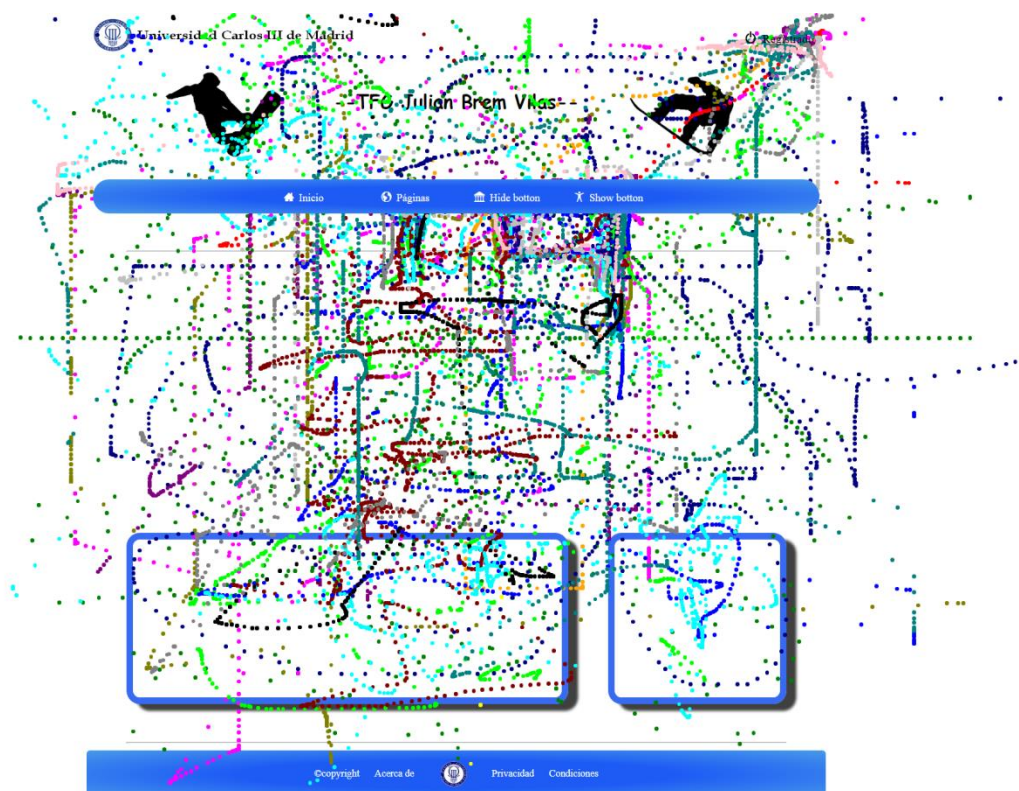


Ilustración 4.7. Inicio: Usuarios registrados

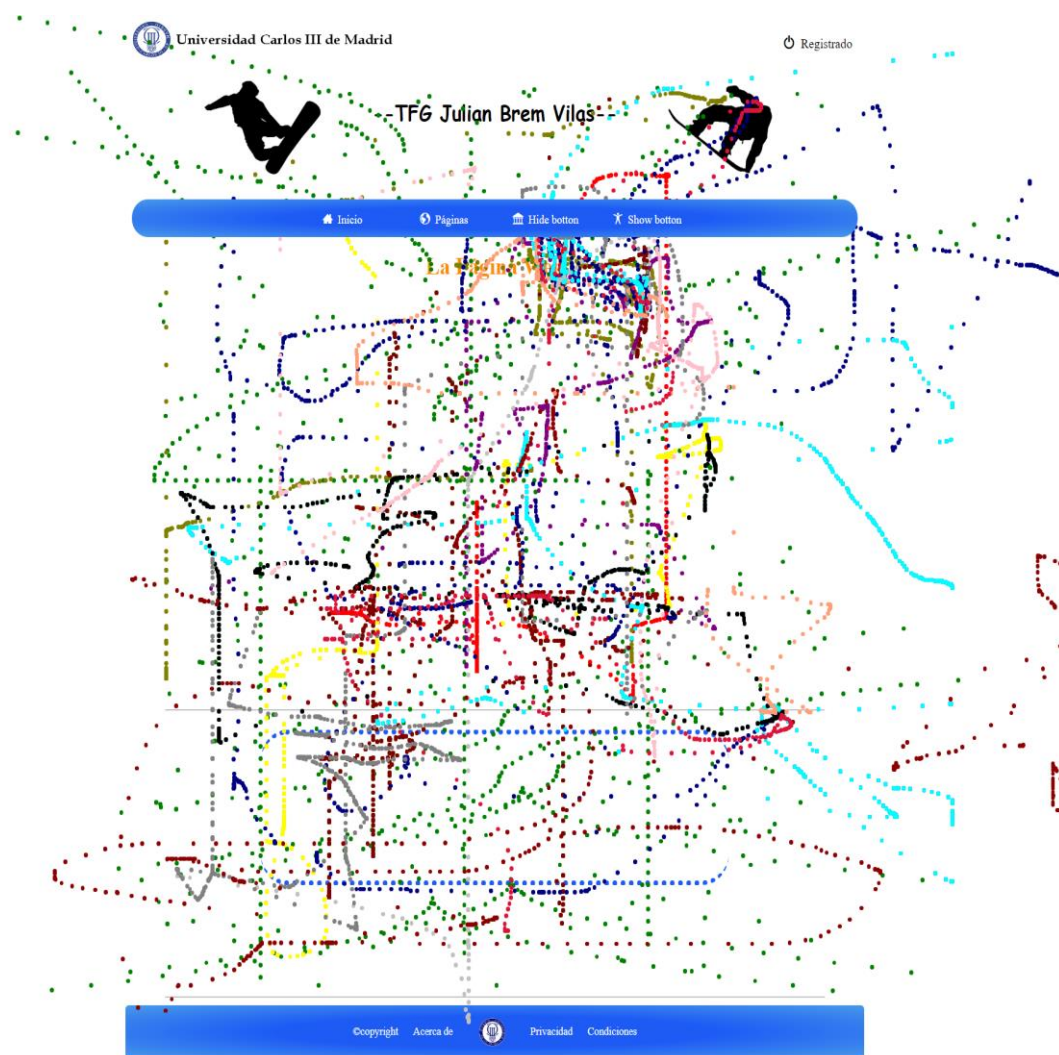


Ilustración 4.8. Página web: Usuarios registrados

Normalmente los robots o programas reaccionan frente a un estímulo o un objetivo que ha sido previamente configurado, por lo tanto nuestra hipótesis inicial a la hora de analizar los datos es que los seres humanos al navegar por internet van a tener un comportamiento más aleatorio, es decir sus movimientos (coordenadas del ratón) no van a seguir un patrón determinado, sin embargo los de un robot van a ser más lineales, ordenados y estarán enfocados a objetivos claros.

Como se puede apreciar en las diferentes imágenes, los movimientos de los usuarios son impredecibles y se concentran sobre todo en zonas dónde se puede interactuar, como la barra del menú de navegación, los botones e imágenes que contienen algún efecto o link. Por otra parte se puede apreciar que la gran mayoría de los movimientos se han realizado en zonas donde hay contenido y una información que puede ser interesante para el usuario. Así mismo podemos ver que las líneas que contienen menos puntos se

deben, a que el desplazamiento ha sido más rápido que en el caso de las líneas que contienen más puntos.

El análisis de los tiempos de visualización nos da una pista sobre el posible usuario, ya que la curiosidad o el interés por el contenido de cada página determinan el tiempo de visita de cada una de ellas en el caso del ser humano.

Para poder ver este aspecto, más en detalle, hemos cronometrado a un total de 10 usuarios visualizando alguna de nuestras páginas web. Para poder tener unos tiempos homogéneos hemos puesto como requisito leer el contenido al completo de cada una de las páginas que se vayan a estudiar y se registrará de forma individual el tiempo de cada página.

TIEMPOS DE VISUALIZACIÓN DE LOS USUARIOS EN SEGUNDOS

Usuarios	Nombre de las sub-páginas			
	Inicio	Página web	Lesiones	Estiramientos
1	43,42	120,45	200,61	161,47
2	45,55	138,22	204,59	158,89
3	42,75	121,12	192,37	140,26
4	45,30	132,40	204,73	168,28
5	46,88	122,63	206,76	160,33
6	44,49	125,87	207,12	164,89
7	47,25	132,35	205,93	159,28
8	51,31	141,32	212,25	166,76
9	49,46	139,24	211,46	170,57
10	45,78	136,68	205,19	163,39
Promedio	46,22	131,03	205,10	161,41

Tabla 4.1. Tiempos de visualización de los usuarios

TIEMPOS DE VISUALIZACIÓN DEL BOT EN SEGUNDOS

Nombre de las sub-páginas			
Inicio	Página web	Lesiones	Estiramientos
68,22	75,88	76,27	85,87

Tabla 4.2. Tiempos de visualización del autómata

Como se puede observar en las diferentes tablas, los tiempos de los seres humanos se mantienen dentro de un rango determinado, ya que a diferencia del robot, nosotros nos

detenemos en diferentes puntos de la página, para poder ver y leer todo el contenido que nos ofrece.

Sin embargo el robot tiene la posibilidad de modificar su tiempo de ejecución variando de esta manera su tiempo de visualización de la página. Este fenómeno viene prediseñado por el creador del autómata y como se puede ver el rango de tiempos de nuestro bot se encuentra entre 65 y 90 segundos dependiendo de la página, pero en ningún momento llega a los tres minutos de visualización de la página web “Lesiones” (Ilustración 4.1.).

Por otro lado los “Clicks” que se han llevado a cabo siguen una distribución similar al desplazamiento, es decir son aleatorios y a veces, hay más de uno en la misma posición, eso quiere decir que al intentar activar un script no han acertado y han tenido que volver a pulsar la tecla del ratón, lo cual nos da otra pista del usuario que puede estar detrás de esos movimientos.

Recogemos también algunas imágenes (Ilustración 4.9.- 4.12.) de los movimientos realizados por el otro protagonista de nuestro estudio, el robot.

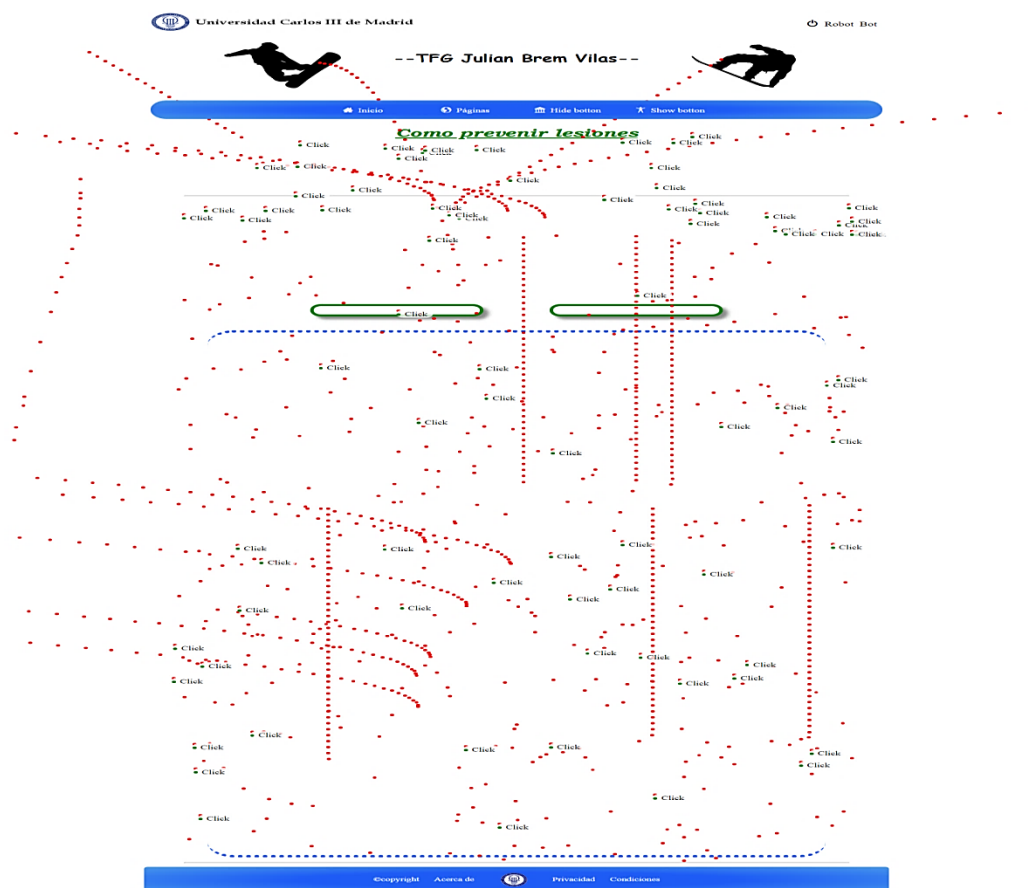


Ilustración 4.9. Estiramientos: Robot

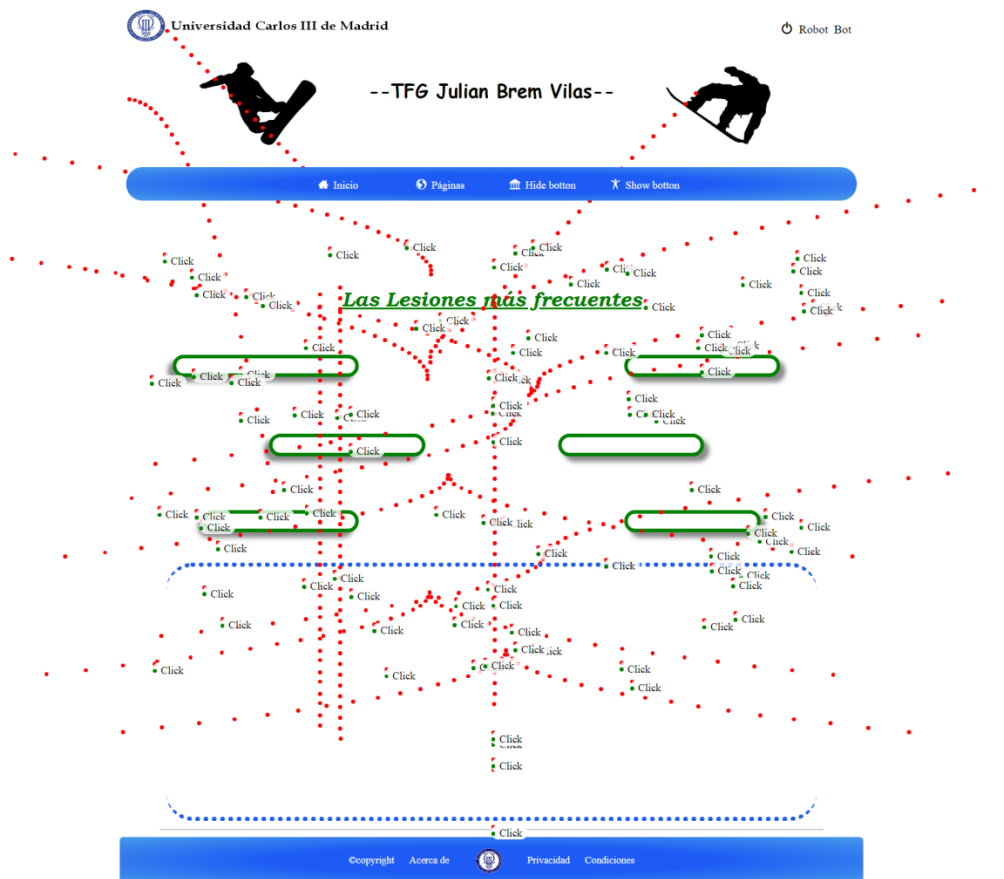


Ilustración 4.10. Lesiones: Robot

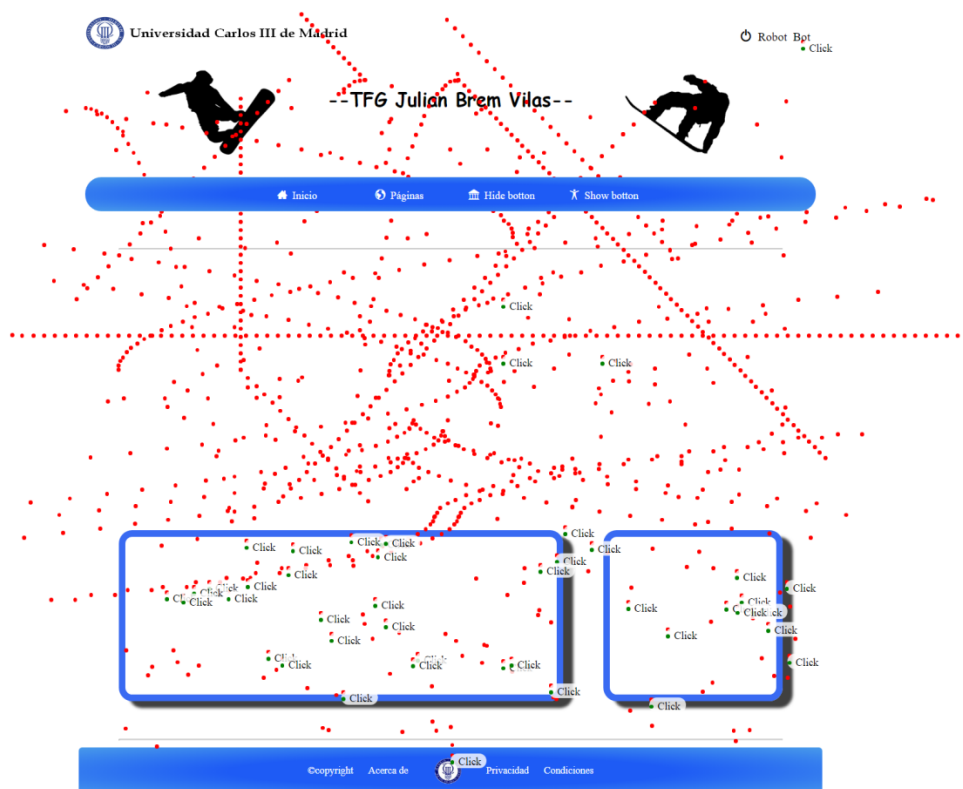


Ilustración 4.11. Inicio: Robot

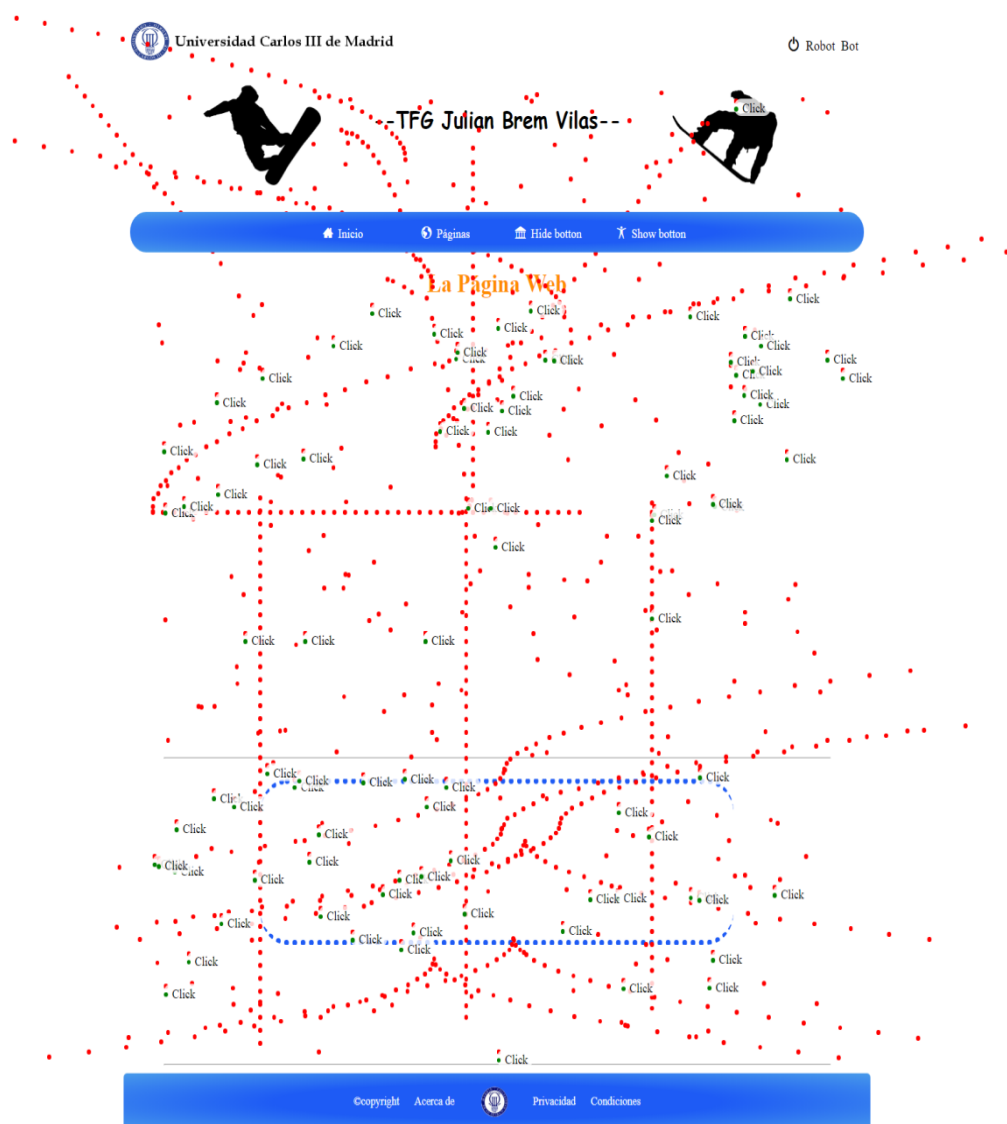


Ilustración 4.12. Página web: Robot

Analizando las imágenes que muestran los movimientos del robot, vemos como el autómatas directamente se dirige a un punto en concreto, sin dejar a penas un rastro por la página (Ilustración 4.9. y 4.12.). Los movimientos son a veces tan rápidos que no podemos capturar los desplazamientos, sin embargo al disminuir la velocidad, podemos observar, que los desplazamientos no son tan fluidos como los de un ser humano (Ilustración 4.11.).

A la hora de hacer “Click” el robot, a diferencia con el ser humano, va directamente al punto deseado y no necesita varios intentos para acertar con el destino. Es más complicado ver varios clicks en un mismo sitio o a distancias mínimas entre ellos. Por otro lado el bot es capaz de posicionarse en diferentes puntos sin tener que dejar una línea que delate su movimiento, es decir puede teletransportarse de un punto a otro y no

dejar rastro y lo mismo sucede con los clicks. Puede generar clicks a gran escala y de forma aleatoria, por lo que engañaría e inutilizaría el estudio (mapas de calor) de algunas herramientas que hemos mencionado en el apartado 2 (Estudio de las tecnologías: Click Tale, etc.), ya que no se podría especificar qué información o que sitio de la página digital ha sido más frecuentada según el número y la disposición de los clicks.

Otro aspecto fundamental son las líneas de movimiento. Comparando un usuario con el bot, el bot tiene movimientos lineales generando rectas, curvas, por el contrario en la del usuario los movimientos son más fluidos e imprevisibles (1 en la Ilustración 4.13.). Por otro lado podemos ver en la diferentes imágenes que los clicks de los seres humanos están vinculados con el contenido de la página, en cambio como bien hemos dicho anteriormente el bot, los clicks que realiza no le mueve la inquietud por conocer la información que le pueda aportar cada uno de los apartados incluidos en la página (2 en la Ilustración 4.13.).

Además podemos ver en la Ilustración 4.10., que el bot, a diferencia del usuario, aparece y desaparece sin dejar huella de su desplazamiento o lo que es lo mismo no existe un movimiento de transición entre los diferentes destinos.

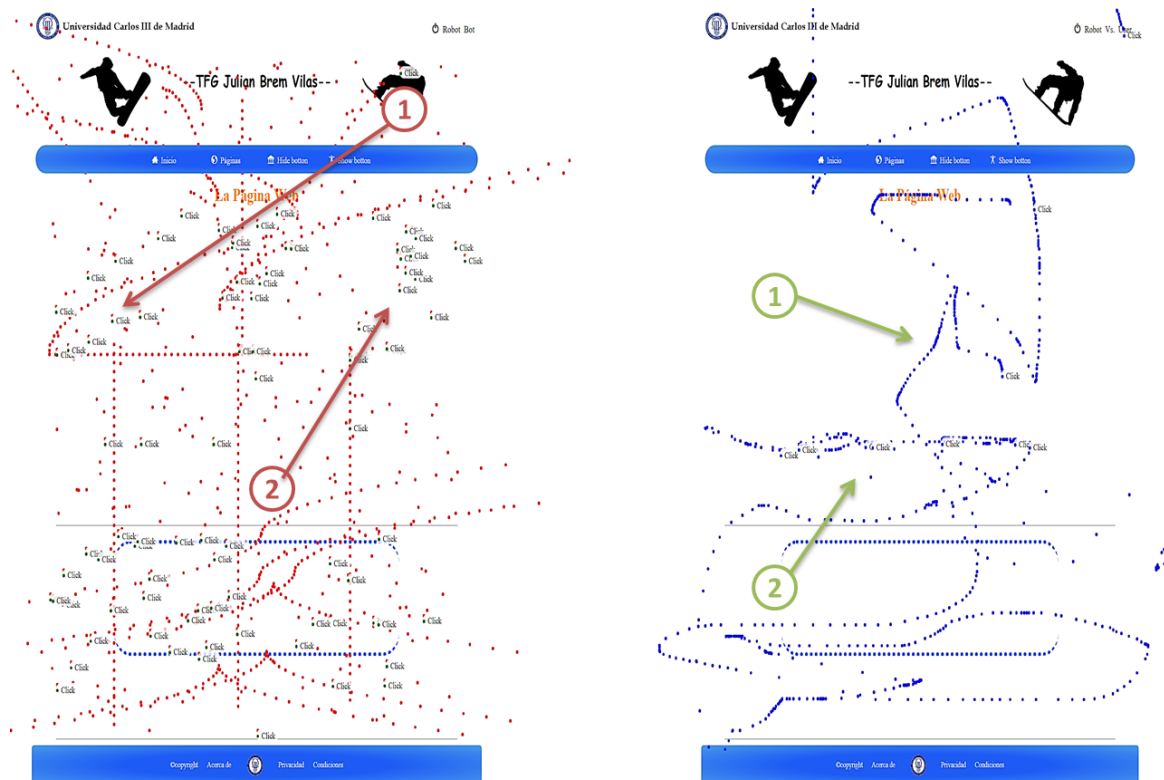


Ilustración 4.13. Diferencias entre ser humano y robot

5. MARCO REGULADOR

Al tratarse de datos vinculantes a los usuarios nos tenemos que adaptar a la directiva sobre la protección y el tratamiento de la información personal de la Unión Europea.

5.1. Directiva Europea de protección de datos.

Esta herramienta para el análisis de los usuarios se encuentra dentro del marco regulatorio europeo⁽¹⁹⁾ según la Directiva 95/46/CE, la cual define entre otras cosas, que se aplicará dicha directiva a cualquier forma de tratamiento de la información personal que vaya a ser recogida y almacenada en ficheros.

Antes de empezar debemos tener claros algunos conceptos⁽²⁰⁾ como son:

- Datos personales: aquella información relacionada con una persona física que pueda ser capaz de identificarla de manera directa o indirecta basándose en su aspecto físico, económico, cultural o social.
- Tratamiento de la información personal: cualquier conjunto de acciones que se lleven a cabo sobre los datos personales de algún individuo, ya sea de forma automatizada o no, incluyendo la recogida de los mismos en una base de datos.

Partiendo de esta base podemos ver que nuestra herramienta se ve afectada por estas dos definiciones, ya que realizará un conjunto de operaciones sobre los datos recogidos de nuestros usuarios, sin embargo no todos los datos que se recogen son relevantes para la posible identificación del individuo (Movimientos del ratón y uso del teclado).

No obstante nuestra herramienta archivará los diferentes datos en ficheros diferentes, utilizando como identificador, dependiendo del tipo de usuario, la dirección IP y el User Agent o el nickname, por ese motivo nuestro programa tendrá que implementarse siguiendo la nueva ley GDPR (General Data Protection Regulation), la cual entró en vigor el 25 de Mayo de 2018 y en la cual se contempla que la dirección IP o el nickname es un dato personal que permite identificar a la persona que está utilizando ese ordenador. Por ello, para cumplir con esta ley, dichos datos se deben anonimizar aplicando una función hash, de forma que no se puede identificar al usuario y sigan sirviendo de identificador unico.

Así mismo la gran mayoría de páginas web suelen trabajar con formularios, Cookies, etc. que contienen información personal de cada usuario por lo que el conjunto de herramientas, que se van a implementar en una web, tienen que ser capaces de adaptarse al reglamento.

Además de la directiva europea existe una ley complementaria específica para la privacidad online. Se conoce como ePrivacy Directive (ePD), la cual viene especificada en la Directiva 2002/58/CE [27], la cual mencionaremos en el siguiente apartado.

5.2. Regulación en España

España sigue las directrices de la Unión Europea y como tal, la ePD se registró en la ley 34/2002, de 11 de julio, en los servicios de la sociedad de la información y de comercio electrónico.

Esta ley se basa en que los usuarios tienen que dar su consentimiento para poder hacer uso de sus datos personales, el cual se consigue a través de un modelo de autorización.

Cuando un usuario da su consentimiento, en España, se debe informar de cómo puede eliminar el almacenamiento de su información personal. El acceso para poder revocar su consentimiento personal según la Agencia Española de Protección de Datos (AGPD) debe estar en cualquier momento accesible a través de la propia página web.

Por otro lado si un usuario no manifiesta expresamente si acepta los términos y condiciones del uso de la información personal y sigue navegando en la página web, se podría entender que este ha dado su consentimiento siempre y cuando se haya informado debidamente del uso del mismo, a través de avisos en la web.

Como todos sabemos las leyes están en continuo cambio, actualizando la información y renovando puntos obsoletos, por lo que hay que estar atento de las nuevas regulaciones, para permanecer al día de los posibles cambio que aparezcan.

6. ENTORNO SOCIO-ECONÓMICO

Este proyecto tiene como finalidad mejorar y ampliar las herramientas ya existentes para la verificación de la autenticidad de los usuarios. Su objetivo es mejorar la veracidad de las estadísticas de vistas de las páginas electrónicas, por lo que está enfocada y diseñada para los dueños de páginas web como también a los publicistas, los cuales pueden pedir que esta herramienta se implante en las plataformas dónde pretenden invertir para poder ver que los usuarios no son falsos.

6.1. Impacto económico

Desde el punto de vista económico esta herramienta puede ser muy interesante sobre todo en las páginas web que se dedican a vender productos o para las redes sociales, ya que pretende visualizar de una forma alternativa la autenticidad de los usuarios que acceden a ellas, por lo cual se podrían descartar muchas visualizaciones fantasmas que alteran las estadísticas de visita.

En el caso de las páginas que ofrecen bienes podría llegar a mejorar los flujos de material (tanto recursos como productos finales) los cuales se ven influenciados, en gran parte, por dichas estadísticas y de esta forma se podrían reducir los costes de producción, almacenamiento, etc.

Además afecta directamente a uno de los motores que mueve más cantidad de dinero a diario en internet, que es, la publicidad online⁽²¹⁾. Este tipo de publicidad se ve afectada directamente por la presencia de bots, ya que en muchas ocasiones los autómatas acceden a los link de publicidad alterando las estadísticas de visita y a menudo acaban otorgando valor a un producto o una publicidad que en realidad no interesa a los usuarios. Por ese motivo como bien vimos anteriormente se han creado algunas herramientas, como es improvely, para poder analizar y descartar las visitas fraudulentas.

Por otro lado las redes sociales son otra gran plataforma que se ven influenciados por el número de visitas o seguidores que reciben a lo largo del día. Este factor es imprescindible para muchas empresas que se dediquen a la recaudación y transmisión de información, ya que su principal objetivo es llegar a mucha gente. Por ese motivo

esta herramienta sería útil implementarla en varias redes sociales para poder verificar el número de usuarios verdaderos y descartar las cuentas o las visualizaciones fantasmas.

6.2. Impacto social

Desde el punto social es una herramienta muy útil que tendrá una gran acogida debido a las posibilidades de mejora que genera en las páginas web. De todo es sabido que es un tema que está a la orden del día debido a que cada vez, el uso de internet es más universal y con mi producto se podrá definir al usuario, lo que será muy interesante para las empresas o incluso para las redes sociales, ya que se basan en los usuarios y podrán eliminar las cuentas falsas.

Los usuarios no verán gran utilidad a esta herramienta, sin embargo a modo de curiosidad podrán ver su comportamiento dentro de la web y visualizar como se han ido desplazando por la página digital como modo de entretenimiento.

7. PLANIFICACIÓN DEL PROYECTO

Todos los proyectos necesitan de una planificación previa y un presupuesto asociado. Bien es sabido que esta es la parte más crítica de la idea, ya que nos va a decir y a enumerar todo lo necesario para poder implantarla de una forma sencilla y a ser posible económica.

7.1. Planificación

A continuación vamos a mostrar un diagrama de GANTT⁽²¹⁾ que nos va a mostrar cómo se ordenan las tareas.

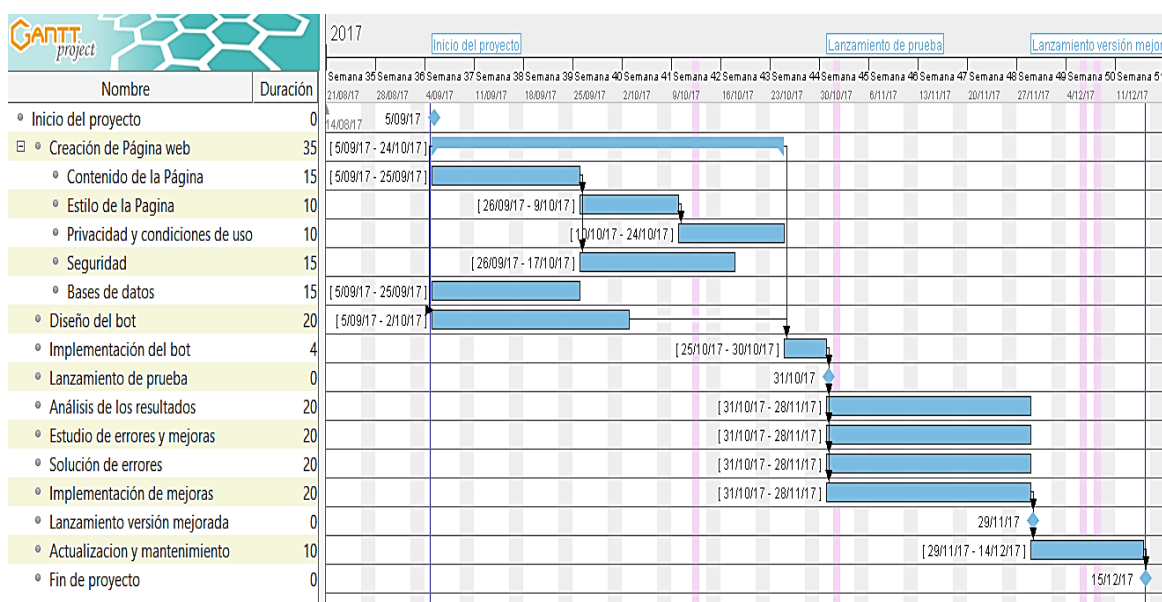


Ilustración 7.1.1. Diagrama de Gantt

El diagrama nos muestra como la duración del proyecto comprende un total de 4 meses. Empezaría el 4 de Septiembre de 2017 hasta el 3 de enero de 2018. La Jornada de trabajo es de 5 horas, excluyendo los fines de semana y las fiestas.

La duración total del proyecto estaría en torno a 70 días con una duración de 350 horas, de las cuales sólo 220 horas estarían dedicadas a la elaboración de esta herramienta.

Los costes siempre van a variar dependiendo de los recursos que se necesiten. En este ejemplo nos hemos basado en la utilización de un estudio con dos ordenadores, dos trabajadores y un servidor (base de datos). El presupuesto está sujeto a los recursos

tanto materiales como humanos que se quieran utilizar y va a variar mucho dependiendo de la organización inicial que se haya planteado. Nuestros recursos han sido:

Para los recursos materiales se han tenido en cuenta todos los accesorios necesarios para poder poner en marcha esta idea como son: los terminales, las licencias, elementos hardware, etc.

RECURSOS MATERIALES

Recursos Materiales	Presupuestos
Dominio Web	20€
Licencia Sublime Text	66€
2 PC + 2 Licencias	2050€
Servidor	1968€
Total	4104€

Tabla 8.1. Recursos Materiales

Por otro lado los recursos humanos se han basado en dos trabajadores para la elaboración de este proyecto y una persona de actualización y mantenimiento de la misma una vez se finalice el trabajo de creación y de mejora.

RECURSOS HUMANOS

Recursos Humanos	Sueldo/h	Horas	Presupuesto
Jefe de proyecto	50€	100	5000€
Programador	25€	445	11125€
Mantenimiento	15€	50	750€
Total	90€	595	16875€

Tabla 8.2. Recursos Humanos

Con los datos de las tablas, podemos concluir que el presupuesto estimado para realizar este proyecto es de **20979€** para un total de **595** horas de trabajo que corresponden a 70 días laborables.

8. CONCLUSIONES

El proceso de elaboración de este estudio ha sido muy laborioso, debido al uso de varios programas y diferentes lenguajes de programación. Para mí ha supuesto todo un reto haber llegado hasta este punto y pienso que el resultado que nos ofrece este análisis del movimiento no refleja todo el trabajo que hay detrás y sin el cual no habiéramos llegado a estas consideraciones finales.

Este proyecto está basado en una herramienta que se ha diseñado con el fin de verificar la identidad del usuario y ha sido creada para poder diferenciar un autómata de un ser humano. Nuestro estudio se ha basado en un bot que hemos desarrollado para simular los movimientos de una persona y no hemos querido hacer en ningún momento una comparación con otro tipo de autómata creado por otro usuario o para otro proyecto.

Es una herramienta visual que identifica la esencia del usuario analizando visualmente los desplazamientos que se han realizado.

Tras analizar las diferentes visitas de los usuarios en la web hemos podido establecer las diferencias fundamentales que existen entre el humano y nuestro autómata y que quedan recogidas en los siguientes puntos:

- El autómata tiene la capacidad de teletransportarse, es decir de aparecer y desaparecer sin dejar huella.
- El humano al desplazarse de un punto a otro siempre dejará un movimiento de transición.
- El tiempo de ejecución, ya que una persona por muy rápida que vaya, nunca podrá llegar a la velocidad de un programa.
- El número de clicks y la distribución que tengan en la página basándose en el contenido de la misma.
- Los movimientos del autómata no están condicionados por el interés que despierta el contenido de la página, sino que sus movimientos son totalmente aleatorios.

Por todo ello podemos concluir que nuestra herramienta es útil, fácil de implementar, pero por supuesto mejorable. Los factores que según nuestro criterio habría que modificar son:

- El acceso mayoritario en el mundo cotidiano al mundo virtual a través de móviles o tabletas nos obliga a adaptarnos a ese, no uso del ratón, en esas circunstancias y a captar los movimientos de forma diferente.
- Es una herramienta visual por lo que el proceso de identificación es más lento. En un mundo donde el ser humano tiende a ser sustituido por la máquina, para mejorar el rendimiento abaratando el coste, tiene poco interés, ya que el análisis final recae sobre un humano.
- Mejorar la gestión de datos para poder liberar espacio de almacenamiento en el caso de que se determine que el usuario es una persona.

I. ABSTRACT

I.I. Motivation of the Project

Over the last years a noticeable increase in security can be observed, when it comes to verifying the user's information before starting a session in a web page.

These days we observe in many occasions segments, where squares with different images appear and the user has to state in which of these squares a certain image can be seen. Occasionally we observed squares where a group of crossed out letters appears or letters that are difficult to read and the user is asked to repeat this sequence (Captcha).

This type of tools has been created in order to verify is the person that accede the web page in reality is human, due to the fact that currently "bots" are not yet able to overcome these tests.

Taking all this into consideration, I asked myself the following question: "Is there a different way to detect if the user, who accesses my digital page, in reality, is human?"

As we all know, we generate different movements when navigating through a web page, be it, when we use the zoom, when we copy certain fragments, when we make a mouse click or simply when we move the mouse. All of these movements give us valuable data on who is behind the navigator.

At present in many web pages the tool called "cookies" is used. This tool collects small amounts of data which are exchanged between the sender (server) and the receiver (navigator). Their main function is to identify the user of a web page in order to offer the contents that are most adequate or interesting for the user according to his navigation behavior. The navigation behavior has been recorded in these small files. In other words, when we access a web page for the first time, a cookie with a small amount of information is saved. This information will be used the second time the user accesses this page with the idea of offering the personalized visit possible to the user. However, this system is not able to detect if the user in question is a human or a robot.

Trying to resolve this problem inspired me for this project, termination of my bachelor degree. I developed I different type of tool which allows us to register the mouse clicks

and other mouse movements, as well as the clicks on the keyboard. By analyzing these classes of movements it was possible to detect certain patterns that allow us to draw conclusions whether these movement patterns belong to a human or a robot.

I.II. Objectives

The main goal of this project is the creation of a tool that allows the analysis of movements from different users (human beings or robots), based on the activities on the web page, extracting consistent movement patterns.

Before generating this tool, it was necessary to design a web page in order to create a platform that could be visited by different visitors. This web page needed to be linked to a data base that allows us to register all different movement patterns of our users. And finally I created a “bot” with the idea of generating a fake user that simulates the behavior of a robot.

The analysis of all these results will enable to resolve the question: “¿is it a robot or a human?”

I.III. Analysis of results

I carried out a visual analysis of the different movements generated, both by anonymous and registered users and compared their movements with those of the programmed machine.

We could observe differences in the use of the mouse clicks, differences in the duration of the visit, but specially, differences in the type of movement registered during the navigation on the page.

These differences are as follows:

- The robot has the capacity to appear and disappear without leaving a foot print (tele- transfer).
- The human will always need a transition movement while moving from one point to the next.
- The velocity of execution: a human, even if trained, will not be able to move as fast as a robot.
- The number of clicks needed and the clicks’ distribution over the page and its content.
- The robot’s movements are not influenced by the pages’ contents, whereas human curiosity is attracted to special places in the web site.

I.IV. Conclusions

The elaboration of this project has been a rather complex process, as various programs and different programming languages had to be used. For me it has been an important challenge being able to reach to this point and being able to detect and analyze properly these movement patterns. This strong effort might not be visible at first glance.

Taking all these into consideration, you may conclude that the tool is useful, relatively easy to implement, yet there is still room for improvement. The elements that to my understanding should be further adapted are:

- The access to virtual world is more and more done via mobile phones and tablets and no more mouse clicks are any longer necessary. Need to adapt the tool to this new reality.
- For being a visual tool the process to identification is relatively slow (human factor). Nowadays in many processes the human tends to be substituted more and more by the machine, in order to improve output and to reduce cost. As in this case the final evaluation still needs to be done by a human, the practical value is limited.
- Improve the data management in order to free storage capacity, once the user has been identified as human (these data are no longer necessary and only consume capacity).

II. REFERENCIAS

- [1] Herramientas para visualizar el comportamiento del usuario, [online] Disponible en: <https://www.antevenio.com/blog/2016/09/herramientas-para-visualizar-el-comportamiento-de-los-usuarios/>
- [2] Click Tale herramienta para el comportamiento de los usuarios, [online] Disponible en: <http://www.doctormetrics.com/2017/08/30/aun-no-conoces-clicktale/#.WxwUwIr-iMp>
- [3] Hotjar herramienta para el comportamiento de los usuarios, [online] Disponible en: <http://growthhackinglabs.com/hotjar-herramienta-growth-hacking/>
- [4] Crazy Egg herramienta para el comportamiento de los usuarios, [online] Disponible en: <https://www.kanlii.com/analisis-web-y-optimizacion/crazy-egg/>
- [5] Improvely: Optimiza tus campañas de Marketing Online, [online] Disponible en: <http://comprardominioweb.com/analitica-web/improvely>
- [6] ¿Qué son los Captcha y cómo funcionan?, [online] Disponible en: <https://www.mediatrends.es/a/119768/captcha-google-que-son-tipos-como-funciona/>
- [7] Test de Turing, [online] Disponible en: <https://matap.dmae.upm.es/cienciaficcio/DIVULGACION/3/TestTuring.htm>
- [8] Elegir el entorno de desarrollo web, [online] Disponible en: <https://programarfacil.com/podcast/36-como-elegir-el-entorno-de-desarrollo-web/>
- [8] ¿Qué es un IDE y cual elegir?, [online] Disponible en: <http://ayudasprogramacionweb.blogspot.com.es/2011/12/que-es-un-ide-y-cual-elegir.html>
- [8] 6 entornos de programación para programa en varios lenguajes, [online] Disponible en: <https://www.redeszone.net/2017/03/24/conoce-estos-6-entornos-programacion-ide-programar-varios-lenguajes/>
- [8] Los 10 lenguajes de programación más populares, [online] Disponible en: <http://noticias.universia.com.ar/consejos-profesionales/noticia/2016/02/22/1136443/conoce-cuales-lenguajes-programacion-populares.html>
- [9] Descripción general de PHP, [online] Disponible en: <https://www.codejobs.biz/es/blog/2013/11/06/descripcion-general-de-php>
- [10] Programación en Java, [online] Disponible en: https://es.wikibooks.org/wiki/Programaci%C3%B3n_en_Java/Caracter%C3%ADsticas_del_lenguaje
- [11] Descripción general sobre ASP.NET, [online] Disponible en: [https://msdn.microsoft.com/es-es/library/4w3ex9c2\(v=vs.100\).aspx](https://msdn.microsoft.com/es-es/library/4w3ex9c2(v=vs.100).aspx)
- [12] Java vs. Php: Eterno debate, [online] Disponible en: <https://www.facilcloud.com/noticias/java-vs-php-eterno-debate/>
- [13] Las mejores bases de datos en 2017, [online] Disponible en: <https://blog.pandorafms.org/es/mejores-bases-de-datos/>

- [13] Comparación entre MySQL y SQL Server, [online] Disponible en:
<http://www.latindevelopers.com/articulo/diferencias-entre-mysql-y-sql-server/>
- [13] Ventajas e inconvenientes de Microsoft SQL, [online] Disponible en:
<https://www.ventajaseinconvenientes.com/ventajas-e-inconvenientes-de-microsoft-sql-server/>
- [14] Diferencias entre NoSQL y SQL, [online] Disponible en:
<https://blog.pandorafms.org/es/nosql-vs-sql-diferencias-y-cuando-elegir-cada-una/>
- [15] ¿Qué es HTML?, [online] Disponible en:
<https://codigofacilito.com/articulos/que-es-html>
- [16] ¿Qué es CSS?, [online] Disponible en:
https://developer.mozilla.org/es/docs/Learn/CSS/Introduction_to_CSS/Como_funciona_CSS
- [17] JavaScript y jQuery, [online] Disponible en:
<https://openclassrooms.com/courses/introduccion-a-jquery/que-son-javascript-y-jquery>
- [18] GDPR (Reglamento general de protección de datos) en España, [online] Disponible en:
<https://acumbamail.com/gdpr/>
- [19] Eur-lex.europa.eu. (2016) Reglamento Europeo, [online] Disponible en:
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679>
- [20] Definición y clases de datos personales, [online] Disponible en:
<http://www.habeasdat.com/faq.html>
- [21] Golpe al fraude en la publicidad online, [online] Disponible en:
<http://www.expansion.com/economia-digital/innovacion/2018/03/27/5aba274946163fc5218b4645.html>
- [22] Gantt Project, Free Project scheduling and management app, [online] Disponible en:
<https://www.ganttproject.biz/>